Multivariate ordered discrete response models*

Tatiana Komarova† William Matcham‡

March 14, 2023

Abstract

We introduce multivariate ordered discrete response models with general rectangular structures. From the perspective of behavioral economics, these non-lattice models correspond to broad bracketing in decision making, whereas lattice models, which researchers typically estimate in practice, correspond to narrow bracketing. In these models, we specify latent processes as a sum of an index of covariates and an unobserved error, with unobservables for different latent processes potentially correlated. We provide conditions that are sufficient for identification under the independence of errors and covariates and outline an estimation approach. We present simulations and empirical examples, with a particular focus on probit specifications.

Keywords: Ordered response, non-lattice structure, binary decision tree, identification, semiparametric models, broad bracketing, narrow bracketing

JEL Classification: C14, C31, C35, D9

*We are grateful to Elie Tamer and participants at Yale University, Harvard-MIT, Erasmus University Rotterdam, and LSE-STICERD seminars for helpful comments.

†Department of Economics, University of Manchester. tatiana.komarova@manchester.ac.uk.

‡Department of Economics, London School of Economics and Political Science. w.o.matcham@lse.ac.uk.

1 Introduction

Ordered response models are a primary tool for empirical researchers, with applications in many disciplines. In economics, applications range from levels of risk aversion (Malmendier and Nagel, 2011) to political violence (Besley and Persson, 2011), with numerous in between. An important contribution to this literature, Cunha, Heckman, and Navarro (2007) describes several economic applications and provides an extensive coverage of univariate ordered response models as models of rational choice. Some empirical practice considers multiple univariate ordered responses together but implicitly assumes that the decision thresholds (equivalently, decision rules) across separate dimensions are independent.[1] We refer to such designs as *lattice models*, since the nodes formed by intersections of decision thresholds across all dimensions form a lattice structure in the multidimensional space. The left panel of Figure 1 illustrates a lattice model in two dimensions. When responses across several dimensions are determined by a single economic agent, from the perspective of behavioral economics lattice models correspond to a *narrowly bracketing* decision maker. In lattice models, the agent's decision rules in different dimensions are independent.

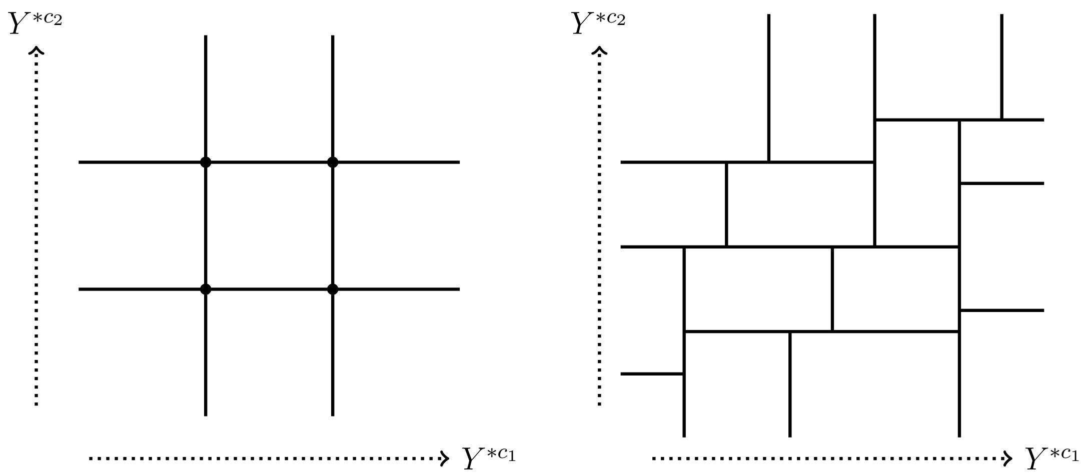
Bracketing effects are central to understanding elements of human choice. However, until recently, the distinction between broad and narrow bracketing has been overlooked in both theoretical and applied economics (Read, Loewenstein, and Rabin, 1999). Traditional economic theory assumes that individuals bracket broadly by maximizing well-defined global utility functions, yet many phenomena are difficult to rationalize if agents bracket decisions this way. For example, the levels of risk aversion required to explain the prices of various forms of insurance seem implausible in magnitude if agents broadly bracket all risks they face (Cicchetti and Dubin, 1994).[2] Furthermore, analysts have turned to models of narrow bracketing to *ex-post* rationalize otherwise hard-to-explain empirical findings.

While individuals may lack the cognitive capacity to analyze multiple relevant choices jointly, it would seem equally unappealing to assume on the other extreme that all decisions are made independently. Resultantly, it is important to have available econometric tools flexible enough to allow for all possibilities of bracketing, letting the data identify the degree of bracketing in different choice dimensions. This is especially true as the more general model of broadly bracketed

[1] Section 2 describes some empirical examples.

[2] In another example, Camerer, Babcock, Loewenstein, and Thaler (1997) finds clear empirical evidence that New York City cab drivers have a *negative* elasticity of hours worked with respect to the daily wage. This phenomenon is hard to rationalize with broadly bracketing, wage maximizing agents. They explain their findings through daily earnings targets, arguing that these workers narrowly bracket work decisions each day at a time.

Figure 1: Models with a lattice (left) and a non-lattice (right) structure



decision making underlies the bedrock assumption of maximization of a global utility function. Yet, no general econometric tools exist in ordered response settings to allow for broad bracketing.

Our central and novel contribution is to introduce and analyze multivariate ordered response models corresponding to *broad bracketing* decision makers, whose decision rules in different dimensions are interdependent. When moving from a single dimension to multiple, researchers often face modeling choices that generate different frameworks of varying complexity.[3] In the context of ordered response, we construct multivariate models that fulfil two desiderata: they should (i) include narrow bracketing designs as a special case, and (ii) preserve prominent features of univariate ordered response models such as threshold-based decisions. In the broad bracketing models we focus on, the nodes formed by intersections of decision thresholds across all the dimensions no longer create a lattice structure. Thus, we refer to these models as *non-lattice*. The right panel of Figure 1 displays an example of a non-lattice design.

In between lattice and general non-lattice are intermediate designs of interest.[4] We focus on the appealing case of *hierarchical models*. These models are generated by a hierarchical decision process where decisions are made sequentially, rather than concurrently. We also show how to describe these models with binary decision trees.

We start our formal analysis by defining non-lattice, lattice, and hierarchical models. In addition to being of stand-alone interest to researchers, hierarchical ordered response models help to

[3]For instance, there are many alternative definitions of a multivariate median, even though the notion of a median in a single dimension is unique.

[4]To continue our analogy, behavioral economics considers *partial narrow bracketing*, which is an intermediate case between broad bracketing and narrow bracketing.

formulate sufficient conditions that guarantee coherency of the more general non-lattice ordered response model.

Following this, we introduce semiparametric specifications of our three models. We model the d^{th} continuous latent process as a sum of an unobservable term ε_d and the index $x_d\beta_d$, which combines observed covariates x_d and an unknown parameter vector β_d. Examples of non-lattice models arising from simultaneous equations follow, with lattice and hierarchical models corresponding to special cases of simultaneity. At this point, we also present a microfoundation of non-lattice models from the perspective of utility maximization.

Our econometric content begins in section 5. We provide formal results on the identification of semiparametric versions of lattice and non-lattice models under the independence of the vector of unobservables $(\varepsilon_1, \ldots, \varepsilon_D)$, collected across all latent processes, from the vector of observables $(x_1, \ldots, x_D)$. Next, we discuss the identification of parametric models when the distribution of joint distribution of errors belongs to a known parametric family. We focus on probit specifications because of their popularity and convenience in modeling dependence across unobservables. Our theoretical content ends with a discussion of estimation in semiparametric and parametric models. A rigorous estimation technique for a general semiparametric non-lattice model is beyond the scope of this particular paper, but we discuss natural directions and explain their relationship to literature on univariate semiparametric ordered response models. On parametric estimation, we provide more detail, including an asymptotic distribution.

Finally, we put our newly-developed parametric estimators to use in simulations and empirical examples. We present Monte Carlo experiments illustrating the deleterious consequences of estimating a misspecified lattice (narrow bracketing) on data generated from non-lattice (broad bracketing) models. The experiments show that when a non-lattice model generates the data, misspecified lattice models estimate significant biases in most parameters. Finally, we give empirical applications that estimate broad bracketing decision making in the context of financial payment choice.

2 Literature review

This paper chiefly contributes to the literature on the economic content of ordered choice models. A leading example is Cunha, Heckman, and Navarro (2007), which examines the economic foundations of ordered discrete choice models. The paper develops a "generalized ordered choice model" to allow for thresholds dependent on observables and unobservables. In doing this, they jointly analyze discrete choices and associated choice outcomes and accommodate uncertainty at the individual level. The model generalizes the standard ordered choice model, which typically has fixed thresholds. The authors develop conditions for nonparametric identification and provide examples of economic models that the generalized ordered choice model can represent. The work also shows that in dynamic contexts, there are restrictions on the arrival of new information and information processing that enables applications of the generalized ordered choice model to dynamic discrete choice such as the choice of schooling years. An earlier example of an ordered choice model with varying thresholds generated from dynamic, sequential choice is Cameron and Heckman (1998). Other papers on ordered choice models with random thresholds include Heckman, Lalonde, and Smith (1999); Carneiro, Hansen, and Heckman (2003); Lewbel (2003). Small (1987) and Bhat and Pulugurta (1998) present alternative microfoundations of ordered choice as random utility maximization and range-based utility maximization, respectively. Finally, Boes and Winkelmann (2006) provides other noteworthy extensions of the traditional univariate ordered response models.

We contribute to this literature by studying the identification and economic content of ordered response models in which thresholds depend on the realization of other endogenous variables, as opposed to regressors and unobservables. This case requires a model of the joint determination of all endogenous variables that influence the thresholds and implies a more flexible structure on thresholds than the one implied by fixed thresholds and univariate stochastic thresholds determined by regressors and errors. Regarding *dynamic* discrete choice, our model is similarly an ex-post representation of a dynamic choice, such as years of schooling. However, it allows for the interaction of multiple *interdependent* dynamic discrete choices. For an example, consider individuals' choices on part-time work *and* education, with interdependence entering not only through correlation of unobservables in latent processes but also through decision rules. In particular, hierarchical models can be microfounded by a dynamic sequence of alternate decisions between outcomes.[5]

A related literature on discrete choice considers strategic interactions, in which outcomes for one player depend on the actions by other players (Tamer, 2003; Berry and Reiss, 2007; Ciliberto and Tamer, 2009; Honore and De Paula, 2010; Chesher and Rosen, 2017, 2020; Aradillas-López

[5]The literature on dynamic discrete choice models is closely related to the literature on dynamic treatment effects; see, for example, Heckman and Navarro (2007); Abbring and Heckman (2007).

and Rosen, 2022). Every agent in this framework corresponds to a separate dimension, and best responses often result in incoherency. The current paper does not address or model strategic interaction of several agents. Instead, we consider a single economic agent deciding along several dimensions. By construction, for a logically consistent agent this decision problem is coherent, and as a result, the non-lattice ordered response models we propose are coherent.[6]

We also contribute to the literature on choice bracketing.[7] This literature is mainly theoretical and experimental (Tversky and Kahneman, 1981; Read, Loewenstein, and Rabin, 1999; Thaler, 1999; Rabin and Weizsäcker, 2009; Ellis and Freeman, 2020; Lian, 2020; Camara, 2021; Zhang, 2021), with some descriptive and few structural empirical applications (Camerer, Babcock, Loewenstein, and Thaler, 1997; Thakral and Tô, 2021).[8] We provide an econometric framework in which researchers can estimate the extent of broad versus narrow bracketing and test for broad bracketing in decision-making by jointly testing if the thresholds in the latent space form a lattice model. We apply this test in our empirical example on online payment instruments and can strongly reject the null of narrow bracketing.

Finally, we add to the empirical literature estimating multivariate ordered response models. This literature contains numerous applications and we refer the reader to Greene and Hensher (2010) for a detailed summary, including a review of recent applications of the bivariate ordered probit model.[9] We estimate bivariate ordered choice models with non-lattice structures. Existing applications assume a lattice structure on the threshold space. To give a specific example, Filer and Honig (2005) studies the joint determination of pension characteristics (age at which eligible) and retirement age, with both dependent variables taking one of five discrete values (less than 62, 63, 64, 65, and greater than 65). Their econometric specification implies narrow bracketing of choices on pension characteristics and retirement age, despite the broad bracketing of this decision being a theme of their work.

[6]Coherency was first considered in Heckman (1978). Tamer (2003) was first to distinguish between incoherency and incompleteness in the context of games and simultaneous strategic actions. We do no make such a distinction in our model.

[7]Several different names have been given to this concept, including *sequential* and *simultaneous* choice (Simonson and Winer, 1992); *narrow* and *broad* decision frames (Kahneman and Lovallo, 1993), *local* and *overall* value functions (Heyman, 1996) and *isolated* and *distributed* choice (Herrnstein and Prelec, 1991).

[8]Tversky and Kahneman (1981) describe a classic example of an experiment in which participants display narrow bracketing.

[9]Applications of trivariate ordered probit models include Buliung and Kanaroglou (2007); Genius, Pantzios, and Tzouvelekas (2006); Scott and Kanaroglou (2002)

3 Model

Now, we present a formal description of our three main classes of multivariate ordered response models.

3.1 Definition of non-lattice, lattice, and hierarchical models

We consider a multivariate ordered discrete response model, which describes a decision process for a single agent joint along $D \geq 2$ dimensions. This decision process maps an underlying D-variate latent continuous metric $(Y^{*c_1}, \ldots, Y^{*c_D})$ into a D-variate discrete metric $(Y^{c_1}, \ldots, Y^{c_D})$. Discrete responses in dimension c_d are denoted as $y_j^{(d)}$, $j = 1, \ldots, M_d$, with

$$y_1^{(d)} < \ldots < y_{M_d}^{(d)}.$$

The decision rules mapping the continuous metric into the discrete metric have a general rectangular structure in the latent space. This leads to the definition of the non-lattice model, which is the most general of the three models we present:

Definition 1 (Non-lattice model) *A multivariate ordered discrete response model is a non-lattice model if*

$$(Y^{c_1}, \ldots, Y^{c_D}) = \left(y_{j_1}^{(1)}, \ldots, y_{j_D}^{(D)}\right) \quad \Longleftrightarrow \quad (Y^{*c_1}, \ldots, Y^{*c_D}) \in R_{j_1,\ldots,j_D},$$

where the D-dimensional rectangle $R_{j_1,\ldots,j_D}$ is

$$R_{j_1,\ldots,j_D} \equiv \bigtimes_{d=1}^{D} \left(\alpha^{(d)}_{j_1,\ldots,j_{d-1},j_d-1,j_{d+1},\ldots,j_D}, \alpha^{(d)}_{j_1,\ldots,j_{d-1},j_d,j_{d+1},\ldots,j_D}\right], \tag{1}$$

with natural normalization conditions on the thresholds $\alpha^{(d)}_{j_1,\ldots,j_{d-1},j_d,j_{d+1},\ldots,j_D}$:

$$\forall d = 1, \ldots, D, \quad \alpha^{(d)}_{j_1 \cdots j_d \cdots j_D} = +\infty \;\; \textit{when} \;\; j_d = M_d, \tag{2}$$

$$\alpha^{(d)}_{j_1 \cdots j_d \cdots j_D} = -\infty \;\; \textit{when} \;\; j_d = 0. \tag{3}$$

We call this a *non-lattice* model since the nodes $(\alpha^{(1)}_{j_1,\ldots,j_{d-1},j_d,j_{d+1},\ldots,j_D}, \ldots, \alpha^{(D)}_{j_1,\ldots,j_{d-1},j_d,j_{d+1},\ldots,j_D})$ at the intersection of decision thresholds across all D dimensions *do not* form a lattice in $\mathbb{R}^D$, unlike a special class of these models (fittingly, *lattice* models) that we discuss later. The right panel of Figure 1 depicts a non-lattice structure when $D = 2$.

The vector latent process is subject to randomness, making the issue of coherency relevant. By coherency in terms of observables, we mean the condition that the probabilities of all discrete responses sum to one. By coherency in the latent space we mean the condition that, for a given set of decision thresholds, the non-overlapping rectangles $R_{j_1,\dots,j_D}$ partition the latent D-dimensional space. These two ways at looking at coherency are equivalent if one can take an generic distribution of observables and unobservables. Since we describe a decision process by a logically consistent single agent, it shall always satisfy the coherency condition. Our description thus far, however, has not indicated conditions on $\alpha^{(d)}_{j_1,\dots,j_{d-1},j_d,j_{d+1},\dots,j_D}$ across different indices that guarantee coherency. We will come back to this after we introduce hierarchical models.

Since the non-lattice model represents a decision maker whose decision rule is interdependent across different dimensions, we can think of them as a decision maker who, in the terminology of the behavioral economics, *broadly brackets* (Read, Loewenstein, and Rabin, 1999; Rabin and Weizsäcker, 2009).

In traditional models, each threshold $\alpha^{(d)}_{j_1,\dots,j_{d-1},j_d,j_{d+1},\dots,j_D}$ is taken to be independent of index j_h, for $h \neq d$. In this case, the decision rule in each dimension d is independent of decision rules in other dimensions. Thus, the joint decision rule can be characterized in individual dimensions, which motivates our definition of lattice models that follows.

Definition 2 (Lattice model) *A multivariate ordered discrete response model is a lattice model if*

$$(Y^{c_1}, \dots, Y^{c_D}) = \left(y^{(1)}_{j_1}, \dots, y^{(D)}_{j_D}\right) \quad \Longleftrightarrow \quad Y^{*c_d} \in \mathcal{I}^{(d)}_{j_d} \equiv (\alpha^{(d)}_{j_d-1}, \alpha^{(d)}_{j_d}] \quad \forall d = 1, \dots, D \tag{4}$$

with natural normalization conditions on the thresholds $\alpha^{(d)}_{j_d}$, $j_d = 0, \dots, M_d$*:*

$$\begin{aligned} \forall d = 1, \dots, D, \quad & \alpha^{(d)}_{j_d} = +\infty \;\; \textit{when} \;\; j_d = M_d, \\ & \alpha^{(d)}_{j_d} = -\infty \;\; \textit{when} \;\; j_d = 0. \end{aligned}$$

Lattice models correspond to a decision maker who *narrowly brackets*, since decisions are made dimension-by-dimension, as opposed to jointly. We refer to such models as *lattice models* since the nodes $(\alpha^{(1)}_{j_1}, \ldots, \alpha^{(D)}_{j_D})$ form a lattice in $\mathbb{R}^D$. These models are automatically coherent without any further restrictions on thresholds and are nested in the class of non-lattice models. Figure 1 gives an example of a lattice structure for $D = 2$. Lattice models are easier to estimate than non-lattice models, but will misspecify a decision maker who broadly brackets.

Importantly, when we talk about decision makers who broadly or narrowly bracket, we refer to, respectively, interdependence or independence of *decision rules*, fully captured by the *thresholds*. Decisions themselves can be correlated in both lattice and non-lattice models because of the correlation in underlying latent processes Y^{*c_d}, $d = 1, \ldots, D$, even after conditioning on observables. Thus, one of the main identification challenges in non-lattice models is to separate the correlation in unobservables from the interdependence of decision rules.

There are several intermediate cases between lattice and non-lattice decision models, and one of particular appeal is the class of *hierarchical models*. One can think of a hierarchical decision process as a process where decisions are made sequentially. The sequential nature of decision making may be due to the agent's preference of doing so or because of the sequential arrival of information. We formally define hierarchical models recursively.

Definition 3 (Recursive definition of a hierarchical model) *A multivariate ordered discrete response model is hierarchical if $M_d = 1$ for all $d = 1, \ldots, D$, or there exists $d_1 \in \{1, \ldots, D\}$ and $j_1(d_1)$ such that*

1. $Y^{c_{d_1}} > y^{(d_1)}_{j_1(d_1)} \iff Y^{*c_{d_1}} > \alpha^{d_1}_{j_1(d_1)}$
2. *The sub-model defined conditional on $Y^{c_{d_1}} > y^{(d_1)}_{j_1(d_1)}$ (equivalently, conditional on $Y^{*c_{d_1}} > \alpha^{d_1}_{j_1(d_1)}$) is hierarchical.*
3. *the sub-model defined conditional on $Y^{c_{d_1}} \leq y^{(d_1)}_{j_1(d_1))}$ (equivalently, conditional on $Y^{*c_{d_1}} \leq \alpha^{d_1}_{j_1(d_1)}$), is hierarchical.*

Hierarchical models are represented as a binary decision tree. Figure 2 depicts a bivariate hierarchical decision process and the binary decision tree in Figure 3 represents the hierarchical process in Figure 2.[10]

Hierarchical models are coherent by definition, since we have a partition of the latent space at each level of the decision tree. These models aid in formulating the coherency condition in the

[10]Such a binary decision tree representation need not be unique.

Figure 2: Model with a hierarchical decision structure

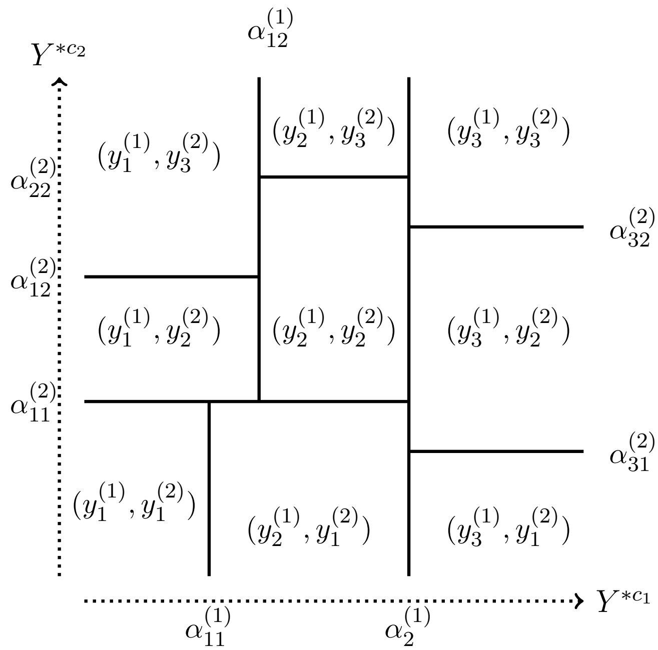
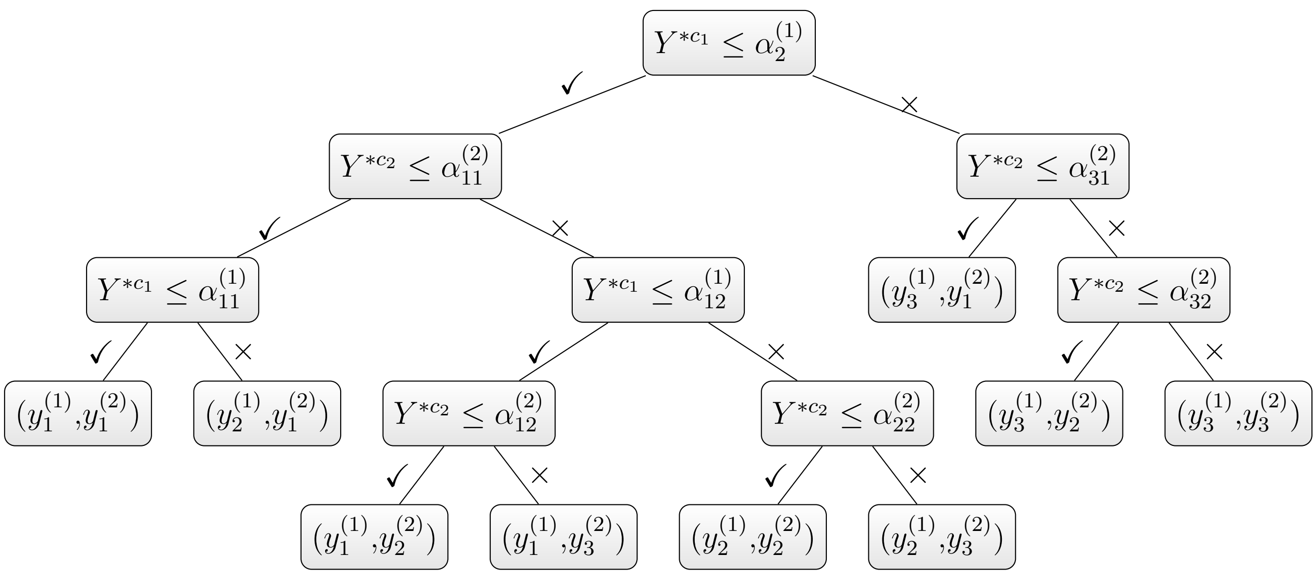


Figure 3: Binary decision tree for the hierarchical model in Figure 2

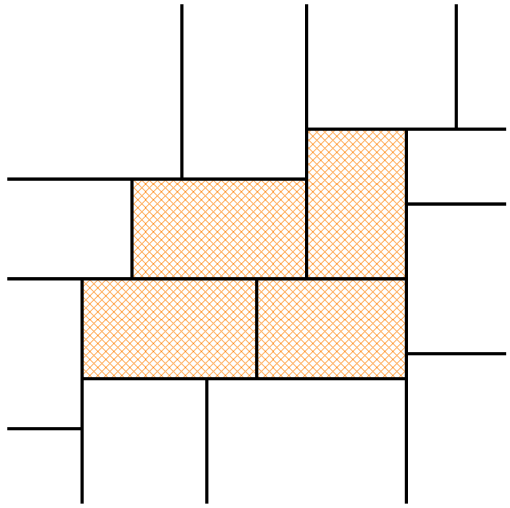


general non-lattice model. To explain this, first we introduce local decision models and local hierarchical models.

Definition 4 (Local decision model) *A local decision model is a model of discrete response* *<u>conditional</u> on discrete responses being among one of 2^D adjacent responses*

$$(y^{(1)}_{j_1+\ell_1}, y^{(2)}_{j_2+\ell_2}, \ldots, y^{(D)}_{j_D+\ell_D}), \quad \ell_d \in \{0, 1\}, \quad d = 1, \ldots, D.$$

FIGURE 4: Intuition for a non-lattice model being locally hierarchical when $D = 2$



In this model, the decision maker chooses $(y^{(1)}_{j_1+\ell_1}, y^{(2)}_{j_2+\ell_2}, \ldots, y^{(D)}_{j_D+\ell_D})$ *if the underlying vector latent process falls into rectangle* $R_{j_1+\ell_1, j_2+\ell_2, \ldots, j_D+\ell_D}$ *conditional on this vector latent process being in the region*

$$\bigcup_{d=1}^{D} \bigcup_{\ell_d \in \{0,1\}} R_{j_1+\ell_1, j_2+\ell_2, \ldots, j_D+\ell_D}.$$

Definition 5 (Locally hierarchical model) *A model is locally hierarchical if each of its local decision models are hierarchical.*

We can establish that being *locally hierarchical* with the given set of thresholds is sufficient to guarantee that a non-lattice model is coherent. In the bivariate case, it is also necessary for coherency. We show these results in Appendix A along with a detailed analysis of the trivariate case $D = 3$. Here we note that in the bivariate case, coherency (E locally hierarchical) means that

$$\left(\alpha^{(1)}_{j_1+1, j_2} - \alpha^{(1)}_{j_1, j_2}\right) \cdot \left(\alpha^{(2)}_{j_1, j_2+1} - \alpha^{(2)}_{j_1, j_2}\right) = 0 \tag{5}$$

for any pair of indices (j_1, j_2), so that $\alpha^{(1)}_{j_1+1,j_2} = \alpha^{(1)}_{j_1,j_2}$ or $\alpha^{(2)}_{j_1,j_2+1} = \alpha^{(2)}_{j_1,j_2}$. In Figure 4, the dashed region $\bigcup_{d=1}^{2} \bigcup_{\ell_d \in \{0,1\}} R_{j_1+\ell_1, j_2+\ell_2}$, formed by four joined rectangles where each rectangle borders with the other three rectangles, represents a local decision model. In that region, the decision model is hierarchical.

3.2 Semiparametric specification

We write each d^{th} continuous latent process as an index in terms of observable covariates x_d (row vector), unknown parameter β_d (column vector) and an additive unobservable error term ε_d:

$$Y^{*c_d} = x_d\beta_d + \varepsilon_d, \quad d = 1, \dots, D. \tag{6}$$

Notably, the terms in $(\varepsilon_1, \dots, \varepsilon_D)$ can be dependent, which allows for the latent processes Y^{*c_d} to be correlated with each other, even conditional on observable covariates.

4 Examples

Our preferred interpretation of non-lattice and lattice models is broad and narrow bracketing respectively. However, other environments also give rise to non-lattice and lattice structures. We discuss some examples in sections 4.1 and 4.2, and provide additional examples including (i) selection in insurance markets, (ii) advertisement spillover effects, and (iii) financial transfers and distress in appendix D.

4.1 Preferences

Univariate model. To relate our interpretation of a general non-lattice model to the existing tradition in economics and econometrics, we first review the univariate case and note that univariate ordered response models are generated by agents with single-peaked preferences. Indeed, consider a model with discrete responses $y_1 < y_2 < \dots < y_M$ and an agent with the realization (x, ε). They have the *ordinal preferences* given by

$$y_{m^*} \succ y_{m^*-1} \succ \dots \succ y_1, \quad y_{m^*} \succ y_{m^*+1} \succ \dots \succ y_M,$$

where m^* is such that $\alpha_{m^*-1} < x\beta + \varepsilon \leq \alpha_{m^*}$ for a given sequence $\alpha_0 = -\infty < \alpha_1 < \dots < \alpha_{M-1} < \alpha_M = +\infty$. Given the interval nature of the responses, the starting point for a utility function corresponding to such ordinal preferences over y_m would be $\min\{-(\alpha_{m-1} - x\beta - \varepsilon), \alpha_m - x\beta - \varepsilon\}$. Using the "min" functional form on its own however creates indifference (which violates single-peakedness) when $x\beta + \varepsilon$ coincides with one of the thresholds. To resolve such ties, we can take the cardinal utility from choosing y_m as

$$U_m = \min\{-(\alpha_{m-1} - x\beta - \varepsilon), \alpha_m - x\beta - \varepsilon\} \cdot 1(x\beta + \varepsilon \neq \alpha_m) + \Delta \cdot 1(x\beta + \varepsilon = \alpha_m),$$

where $\Delta > 0$ can be arbitrarily small and serves as nothing more than a tie-breaking device. Towards showing the single-peakedness property, consider m^* such that $\alpha_{m^*-1} < x\beta + \varepsilon < \alpha_{m^*}$. Then $U_{m^*} > 0$ and $U_{m'} < 0$ for all $m' \neq m^*$. Moreover, $U_{m_1} > U_{m_2}$ for any $m^* \leq m_1 < m_2$ and $U_{m_1} < U_{m_2}$ for any $m_1 < m_2 \leq m^*$. Finally, in the case of a tie, wherein $x\beta + \varepsilon = \alpha_{m^*}$,

$$U_{m^*} = \Delta > \underbrace{U_{m^*-1} > U_{m^*-2} > \ldots > U_1}_{<0}, \quad \text{and} \quad U_{m^*} = \Delta > U_{m^*+1} = 0 > \underbrace{U_{m^*+2} > \ldots > U_M}_{<0}$$

thereby proving the single-peakedness property.[11]

Multivariate model For illustrative simplicity, consider the bivariate case. We construct a utility function that is maximized uniquely when $(x_1\beta_1 + \varepsilon_1, x_2\beta_2 + \varepsilon_2) \in R_{j_1^*, j_2^*} = \left(\alpha^{(1)}_{j_1^*-1, j_2^*}, \alpha^{(1)}_{j_1^*, j_2^*}\right] \times \left(\alpha^{(2)}_{j_1^*, j_2^*-1}, \alpha^{(2)}_{j_1^*, j_2^*}\right]$, corresponding to the choice of $\left(y^{(1)}_{j_1^*}, y^{(2)}_{j_2^*}\right)$. We understand single-peakedness in multiple dimensions as single-peakedness patterns in each direction. To explain this, denote the utility from choosing $\left(y^{(1)}_{j_1}, y^{(2)}_{j_2}\right)$ as U_{j_1,j_2}. Single-peakedness implies the following two conditions:

1. There is (j_1^*, j_2^*) such that $U_{j_1^*, j_2^*} > U_{j_1, j_2}$ when $j_1 \neq j_1^*$ or $j_2 \neq j_2^*$.
2. For any $(j_1, j_2) \neq (j_1^*, j_2^*)$, let (ℓ_1, ℓ_2) be a direction that moves (j_1, j_2) further away from (j_1^*, j_2^*).[12] Then $U_{j_1, j_2} \geq U_{j_1+\ell_1, j_2+\ell_2}$.

Consider the functional form given by

$$\begin{aligned} U_{j_1,j_2} = {} & \min\{-(\alpha^{(1)}_{j_1-1,j_2} - x_1\beta_1 - \varepsilon_1), -(\alpha^{(2)}_{j_1,j_2-1} - x_2\beta_2 - \varepsilon_2), \alpha^{(1)}_{j_1,j_2} - x_1\beta_1 - \varepsilon_1, \alpha^{(2)}_{j_1,j_2} - x_2\beta_2 - \varepsilon_2\} \times \\ & 1\left[(x_1\beta_1 + \varepsilon_1, x_2\beta_2 + \varepsilon_2) \in R^o_{j_1,j_2}\right] \\ & + \Delta \cdot 1\left(x_1\beta_1 + \varepsilon_1 = \alpha^{(1)}_{j_1,j_2}\right) + \Delta \cdot 1\left(x_2\beta_2 + \varepsilon_2 = \alpha^{(2)}_{j_1,j_2}\right) \\ & + \Delta \cdot 1\left(x_1\beta_1 + \varepsilon_1 = \alpha^{(1)}_{j_1,j_2}\right) \cdot 1\left(x_2\beta_2 + \varepsilon_2 = \alpha^{(2)}_{j_1,j_2}\right), \end{aligned}$$

where $R^o_{j_1,j_2}$ denotes the interior of R_{j_1,j_2}. The role of an arbitrarily small $\Delta > 0$ is to provide a

[11] We can make the definition of U_m more general by adding, for example, the same function of covariates to each U_m.

[12] This means either $\ell_1, \ell_2 \in \{0, 1\}$ or $\ell_1, \ell_2 \in \{-1, 0\}$, and $|j_1 + \ell_1 - j_1^*| + |j_2 + \ell_2 - j_2^*| \geq |j_1 - j_1^*| + |j_2 - j_2^*|$.

tie-breaking rule for certain parts of the border of R_{j_1,j_2}. Towards showing that this functional form delivers uniquely maximized preferences over discrete choice pairs $(y^{(1)}_{j_1}, y^{(2)}_{j_2})$, first consider $(x_1\beta_1 + \varepsilon_1, x_2\beta_2 + \varepsilon_2) \in R^o_{j_1^*,j_2^*}$. In this case $U_{j_1^*,j_2^*} > 0$ and $U_{j_1',j_2'} < 0$ for $j_1' \neq j_1^*$ or $j_2' \neq j_2^*$. Second, if $(x_1\beta_1 + \varepsilon_1, x_2\beta_2 + \varepsilon_2) \in R_{j_1^*,j_2^*}$ and $x_1\beta_1 + \varepsilon_1 = \alpha^{(1)}_{j_1,j_2}$ or $x_2\beta_2 + \varepsilon_2 = \alpha^{(2)}_{j_1,j_2}$ or both, then $U_{j_1^*,j_2^*} > 0$ and $U_{j_1',j_2'} \leq 0$. For a lattice structure, such preferences are guaranteed to be single-peaked. For a non-lattice structure they are not necessarily single-peaked even though they are still uniquely maximized.

4.2 Simultaneous equations

Throughout this paper we consider a single decision maker selecting responses across several dimensions. Nevertheless, in some special cases coherent non-lattice models may even arise as a result of strategic interactions. An illustration of that is the simultaneous entry game in Tamer (2003).

Example 1 (Simultaneous entry game in Tamer (2003)) *A small (A) and a large (B) firm can take actions 0/1 (don't enter/enter) and their payoffs are parametrised*

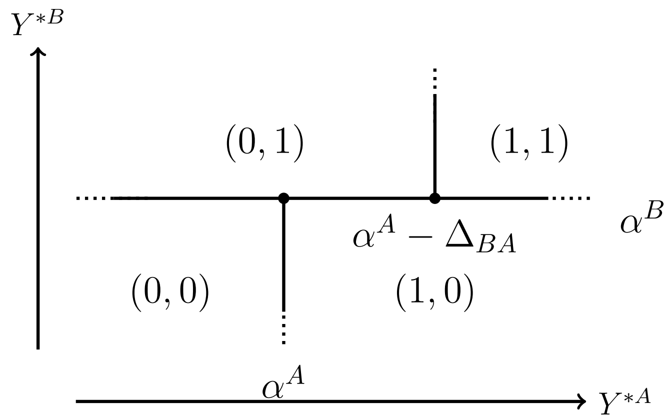
$Y_A \setminus Y_B$	*0*	*1*
0	α^A , α^B	α^A , $x\beta_B + w_B\gamma_B + u_B$
1	$x\beta_A + w_A\gamma_A + u_A$, α^B	$x\beta_A + w_A\gamma_A + u_A + \Delta_{BA}$, $x\beta_B + w_B\gamma_B + u_B + \Delta_{AB}$,

where the presence of the large firm reduces the profit ($\Delta_{BA} < 0$) of the small firm (perhaps the large firm has a large brand advantage), but the presence of the small firm does not affect the profit of the large firm ($\Delta_{AB} = 0$). The discrete responses are then

$$
\begin{aligned}
Y_B &= 1(\overbrace{x\beta_B + w_B\gamma_B + u_B}^{Y^{*B}} > \alpha^B),\\
Y_A &= 1(\underbrace{x\beta_A + w_B\gamma_B + \Delta_{BA}\cdot y_B + u_B}_{Y^{*A}} > \alpha^A).
\end{aligned}
$$

Since $\Delta_{AB} = 0$, we have coherency and completeness (Heckman, 1978; Tamer, 2003) in this model. In the equilibrium, the non-lattice structure shown in Figure 5 represents the decision structure.

FIGURE 5: Latent payoff space for two equations and the non-lattice structure in example 1



Because of the 2×2 nature (two players and two actions) and its *triangular* structure, the strategic interaction problem in Tamer (2003) represents a hierarchical model with the large firm determining the first decision rule and the small firm determining the subsequent decision rules. We can give more general examples of strategic interaction models expressed as simultaneous equations models, resulting in non-lattice models. Based on the univariate model for the choice of differentiated goods in Cunha, Heckman, and Navarro (2007), we provide an example of advertisement spillover effects in Appendix D. Appendix D also has an example (Example 4) of a simultaneous equations model that results in a coherent non-lattice structure.

5 Identification in semiparametric multivariate ordered discrete response models

This section covers the identification of D-variate lattice and non-lattice models from observations on discrete responses and covariates when the latent process in each dimension has the index structure as in (6). We derive identification under either Assumption 1 or the more restrictive Assumption 2, both of which relate to the lack of the statistical relationship between unobservables and covariates.

Assumption 1 *For all $d = 1, \ldots, D$, ε_d is independent of x_d and has a convex support.*

Assumption 2 *The vector of unobservables $(\varepsilon_1, \ldots, \varepsilon_D)$ is independent of $(x_1, \ldots, x_D)$. The support of $(\varepsilon_1, \ldots, \varepsilon_D)$ is a convex set in $\mathbb{R}^D$ with a non-empty interior.*

We employ Assumption 1 in lattice models and Assumption 2 in non-lattice models. In univariate ordered response models, the assumption of independence between the unobservable and covariates is common, being used in Klein and Sherman (2002), Coppejans (2007), and Agresti (2010) among many others.[13]

At this point, we introduce some notation.

Notation 1 *Let $x = (x_1, \ldots, x_D)$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_D)$. Denote the joint c.d.f. of ε as F and the marginal c.d.f. of ε_d as F_d, $d = 1, \ldots, D$. The length of vector x_d is k_d, $d = 1, \ldots, D$. Let $\mathcal{X}_d$ denote the support of x_d and for each d, $d = 1, \ldots, D$, define*

$$S^{(d)} = \left\{ x_d \in \mathcal{X}_d \,|\, \exists\, j_d = 1, \ldots, M_d \;\; \textit{such that } P\left(Y^{(d)} \leq y^{(d)}_{j_d} | x_d\right) \in (0,1) \right\}$$

and

$$S^{(d;j)} = \left\{ x_d \in \mathcal{X}_d \,|\;\; \textit{such that } P\left(Y^{(d)} \leq y^{(d)}_{j} | x_d\right) \in (0,1) \right\}.$$

Let $x_{d,m}$ denote the m^{th} component of x_d. The subvector of x_d including all the components of x_d with the exception of the m^{th} component is denoted $x_{d,-m}$. The term $x_{d,\underline{\ell}:\overline{\ell}}$ denotes the subvector of x_d that includes all the components from $\underline{\ell}$ to $\overline{\ell}$ inclusively, where $\overline{\ell} > \underline{\ell}$. We use analogous notations for β.

Finally, $S^{(d)}_m$ denotes the projection of $S^{(d)}$ on $x_{d,m}$ and $S^{(d)}_{-m}$ denotes the projection of $S^{(d)}$ on $x_{d,-m}$. We use analogous notations for $S^{(d;j)}$.

5.1 Models with lattice structures

We start with identification results for a model with a lattice structure, which is the specific class of non-lattice models satisfying

$$\alpha^{(d)}{}_{j_1,\ldots,j_{d-1},j_d,j_{d+1},\ldots,j_D} = \alpha^{(d)}_{j_d} \quad \forall\, j_1, \ldots, j_{d-1}, j_{d+1}, \ldots, j_D. \tag{7}$$

[13]Some papers (see e.g. Chen and Khan (2003)) on univariate ordered response allow for heteroskedasticity. In our framework, this would correspond to $\sigma_d(x_d, \theta_0)\varepsilon_d$ with independent ε_d. Some other papers further deviate from the setting of independence. Lee (1992) considers ordered response under the median independence assumption from Manski (1975, 1985). In a recent paper, Wang and Chen (2022) take a partial identification approach and consider a generalized maximum score estimator when regressors are interval measured. All of these settings are beyond the score of this paper and provide interesting avenues for extensions of our work.

We formulate an analogue of a rank condition in the form of Assumption 3.

Assumption 3 *$S^{(d)}$ is not contained in any proper linear subspace of $\mathbb{R}^{k_d}$ and $P\left(S^{(d)}\right) > 0$.*

This assumption is traditional in the semiparametric literature of discrete response (Manski (1985, 1988) and Horowitz (2010), among many others) and is implied by more stringent conditions on covariates. Intuitively, it guarantees some minimum desirable variation in the covariates. We are now in a position to provide sufficient conditions that ensure the identification of index parameters β_d, our first result. The proof of this result and all others in text are in Appendix B.

Theorem 1 *Consider a D-variate discrete response model with the index structure in (6). Suppose Assumptions 1 and 3 hold and the model has a lattice structure in the sense of condition (7). For $d = 1, \ldots, D$, the parameter β_d is identified if there is a covariate $x_{d,m(d)}$ in x_d such that for $x_{d,-m(d)} \in S^{(d)}_{-m(d)}$ the support of $x_{d,m(d)}|x_{d,-m(d)}$ intersected with $S^{(d)}_{m(d)}$ contains an interval $(\underline{x}_{d,m(d)}, \overline{x}_{d,m(d)})$ with $\underline{x}_{d,m(d)} < \overline{x}_{d,m(d)}$, and it holds that $\beta_{d,m(d)} = 1$.*

Theorem 1 states that in the latent process d, we need at least one covariate with some continuous variation and a non-zero impact in the d^{th} latent process to identify parameter β_d up to normalization. Of course, $\beta_{d,m(d)}$ can be normalized to -1 instead of 1 if the impact of $x_{d,m(d)}$ is negative rather than positive.

In Theorem 1, we provide identification conditions in terms of *individual* dimensions and *individual* latent processes. This is because the lattice structure allows us to look at each individual dimension without having any interference from or interactions with other dimensions. Resultantly, covariates need not be exclusive to a certain latent process and, in particular, covariates with some continuous variation in latent processes, as required by Theorem 1, can be common to several (potentially all) processes.

Our next step is to analyze the identification of thresholds. Based on the convex support requirement in Assumption 1, we attain the point identification of differences of thresholds (equivalently, the identification of thresholds up to normalization of one thresholds) if for any j and $j+1$, we can find $x_d \in S^{(d;j)}$ and $\tilde{x}_d \in S^{(d;j+1)}$ such that

$$0 < F_d\left(\alpha_j^{(d)} - x_d\beta_d\right) = F_d\left(\alpha_{j+1}^{(d)} - \tilde{x}_d\beta_d\right) < 1.$$

These conditions are effectively in terms of observable probabilities $P(Y^{(d)} \le y_j^{(d)} \,|\, x_d)$ and $P(Y^{(d)} \le y_{j+1}^{(d)} \,|\, \tilde{x}_d)$ and are equivalent to finding $x_d \in S^{(d;j)}$ and $\tilde{x}_d \in S^{(d;j+1)}$ such that $\alpha_{j+1}^{(d)} - \alpha_j^{(d)} = \tilde{x}_d\beta_d - x_d\beta_d$. Since the differences between consecutive thresholds in each dimension are not known a priori, the most straightforward sufficient conditions demand—in addition to the conditions of Theorem 1—a large support from a regressor with continuous variation in each latent process. We state this formally in Theorem 2.

Theorem 2 *Suppose all the conditions of Theorem 1 hold for a particular dimension d. In addition suppose that for covariate $x_{d,-m(d)}$ from Theorem 1, the support of $x_{d,m(d)}|x_{d,-m(d)}$ intersected with $S_{m(d)}^{(d;j)}$ is*

(i) $\mathbb{R}$ if the support of ε_d is unbounded, or

(ii) a sufficiently large interval if the support of ε_d is bounded.

Then in addition to β_d being identified, the differences $\alpha_{j+1}^{(d)} - \alpha_j^{(d)}$, $j = 1, \ldots, M_d - 1$ are identified.

To summarize, in lattice structures, the identification of index parameters β_d and thresholds $\alpha_j^{(d)}$ is separate across different dimensions. This means that we may be able to identify β_d (resp. $\alpha_j^{(d)}$) without other β_ℓ (resp. $\alpha_\ell^{(d)}$), $\ell \neq d$, being identified. This will not be the case in general non-lattice models.

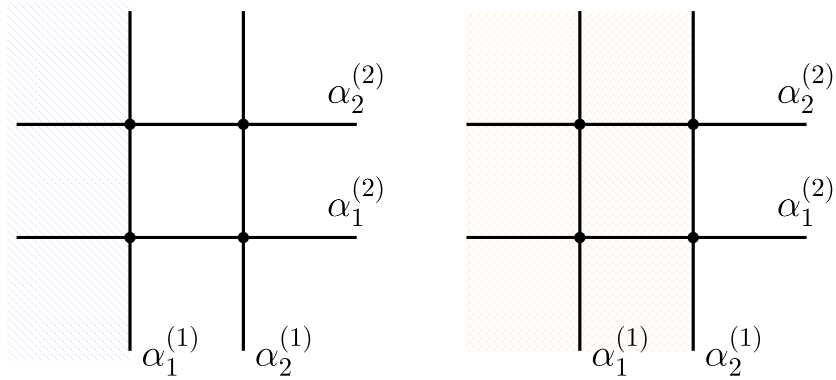
Figure 6, which shows a bivariate lattice model, presents an intuitive summary of the identification strategy in the models with lattice structures. We consider each dimension individually and, within that dimension, express probabilities of discrete values up to certain points in terms of the marginal c.d.f. of the unobservable in that dimension and the index in that dimension.

The result of Theorem 2 immediately implies conditions for identification of marginal distributions of ε_d, $d = 1, \ldots, D$.

Corollary 1 *Suppose conditions of Theorem 2 hold for some d. Then F_d is identified if either one threshold among $\alpha_{j_d}^{(d)}$, $j_d = 1, \ldots, M_d$, is normalized to a known value, or if there is a normalization of one of the values of c.d.f. F_d, say $F_d(e_{0d}) = c_{0d}$, for some known e_{0d} in the support of ε_d and some known $c_{0d} \in (0, 1)$.*

The proof of Corollary 1 is straightforward and is therefore omitted.

Figure 6: Intuition for lattice model identification



Notes: Left region in the latent space corresponds to $P\left(Y^{(1)} \leq y_1^{(1)}|x_1\right)$. Right region corresponds to $P\left(Y^{(1)} \leq y_2^{(1)}|x_1\right)$.

Remark 1 (Identification of joint c.d.f. F) *The result of Corollary 1 does not guarantee identification of the joint distribution of unobservables, even if the conditions of that Corollary hold for every $d = 1, \ldots, D$. The reason is two-fold. First, Assumption 1 does not give any information about how the vector ε relates to the vector x. Second, if the components in ε are not mutually independent conditional on x, then for the identification of F one would need to consider joint outcomes $\left(Y^{(1)} \leq y_{j_1}^{(1)}, \ldots, Y^{(D)} \leq y_{j_D}^{(D)}\right)$ that result in the vector $(\alpha_{j_1}^{(1)} - x_1\beta_1, \ldots, \alpha_{j_D}^{(D)} - x_D\beta_D)$. The issue is that some (or even all) x_d may not have exclusive covariates in them which potentially makes the vector $(\alpha_{j_1}^{(1)} - x_1\beta_1, \ldots, \alpha_{j_D}^{(D)} - x_D\beta_D)$ take values only in a proper subset of the support of $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_D)$.*

However, the identification of the joint c.d.f. F is possible if Assumption 2 holds and each x_d contains a large-support exclusive covariate with a non-zero impact. Essentially, these are the conditions under which we establish full identification in semiparametric non-lattice models in section 5.2 (see Theorem 5 below).

5.2 Models with non-lattice structures

Next, we establish identification in models with non-lattice structures. Now the decision rules in different dimensions may interact in complicated ways and, therefore, analyzing one dimension at a time will not be fruitful. In Figure 7, we illustrate a region in the latent space of a bivariate non-lattice model that corresponds to the probability $P\left(Y^{(1)} \leq y_1^{(1)}|x\right)$. The probability of that region cannot be expressed in terms of values of the marginal c.d.f. F_1 alone. Non-lattice cases therefore require a different approach to identification. Intuitively, the identification of parameters β_d and the threshold structure in these models should be more demanding on the data, especially given an unknown dependence structure of unobservables. This is indeed the case: in this section, we give sufficient conditions for identification in non-lattice models and these conditions are more stringent than those for lattice models.

Identification proceeds in several steps as follows:

1st step: identification of parameters corresponding to exclusive covariates in each process (Theorem 3).

2nd step: identification of parameters corresponding to non-exclusive covariates (Theorem 4).

3rd step: identification of the joint c.d.f. (Theorem 5).

4th step: identification of the thresholds (Theorem 6).

For what remains, we need some additional notation.

Notation 2 *Let $\mathcal{X} \subset \mathbb{R}^{\sum_{i=1}^{d} k_i}$ denote the support of x. For $j_d = 1, \ldots, M_d$, define*

$$S_{all}^{(d)}(j_d) \equiv \left\{ x \in \mathcal{X} : \, P\left(Y^{(d)} \leq y_{j_d}^{(d)} | x\right) \in (0,1) \right\}.$$

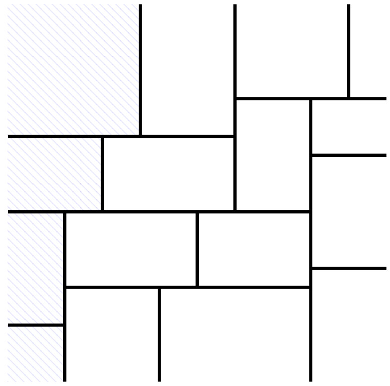
For a subvector z of x, let $S_{all;-z}^{(d)}(j_d)$ denote the projection of $S_{all}^{(d)}(j_d)$ on the subvector obtained by removing z from vector x, and let $S_{all;z}^{(d)}(j_d)$ denote the projection of $S_{all}^{(d)}(j_d)$ on z.

Notation 3 *For each $d = 1, \ldots, D$, let $x_{d,1:L_d}$ denote the subvector of x_d that consists of all the covariates in x_d that are exclusive to the process Y^{*c_d}. Being exclusive means that the conditional distribution $x_{d,i} | x_{-d}$ has a non-degenerate distribution almost everywhere for $x_{-d} = (x_1, \ldots, x_{d-1}, x_{d+1}, \ldots, x_D)$.*

Our final assumption gives an analogue of the rank condition that used in Theorems 3 - 6 below.

Assumption 4 *For any $d = 1, \ldots, D$, there is $j_d = 1, \ldots, M_d - 1$, such that the set $S_{all}^{(d)}(j_d)$ is not contained in any proper linear subspace of $\mathbb{R}^{\sum_{i=1}^{d} k_i}$ and $P\left(S_{all}^{(d)}(j_d)\right) > 0$.*

Figure 7: Region in the latent space for a non-lattice model



Notes: Illustration of region in the latent space corresponding to $P\left(Y^{(1)} \leq y_1^{(1)}|x\right)$ in a bivariate non-lattice model

Theorem 3 gives sufficient conditions for the identification of $\beta_{d;1;L_d}$, $d = 1, \dots, D$, which are the parameters corresponding to the exclusive covariates in each process. It considers, for each d, the marginal probabilities $P\left(Y^{(d)} \leq y_{j_d}^{(d)}|x\right)$ for some (or all) j_d and the region in the D-dimensional space for the continuous latent metric that corresponds to this probability (such as displayed in Figure 7). Even though this region has a complicated structure, we can use the rectangular nature of decision rule cells to express this probability in terms of the joint c.d.f. F of unobservables and indices $x_\ell\beta_\ell$, $\ell = 1, \dots, D$. This probability generally depends on all the indices $x_\ell\beta_\ell$ and is non-increasing with respect to the own index $x_d\beta_d$. Hence, we can map ordinal relations among probabilities $P\left(Y^{(d)} \leq y_{j_d}^{(d)}|x\right)$ to ordinal relations among indices $x_d\beta_d$ provided we keep the values of all the other indices $x_\ell\beta_\ell$, $\ell \neq d$, fixed. From this, identification of parameters corresponding to *exclusive covariates* are obtained subject to some normalization restrictions and to some continuous variation among at least one exclusive covariate in each index.

Theorem 3 *Consider a D-variate discrete response model with the index structure (6). Suppose Assumptions 2 and 4 hold for each $d = 1, \dots, D$, and the model has a coherent (potentially non-lattice) structure.*

Suppose that the following conditions are satisfied:

(a) $L_d \geq 1$ for each $d = 1, \dots, D$ – that is, each process has at least one exclusive covariate.

(b) The coefficient $\beta_{d,1}$ corresponding to $x_{d,1}$ in $x_d\beta_d$ is 1, $d = 1, \dots, D$.

(c) *For each $d = 1, \ldots, D$, there exists j_d that satisfies conditions in Assumption 4 and is such that $S^{(d)}_{all;x_d}(j_d)$ contains a Cartesian product $(\underline{x}_{d,1}, \overline{x}_{d,1}) \times \tilde{S}^{(d)}$, where $\overline{x}_{d,1} > \underline{x}_{d,1}$ and $\tilde{S}^{(d)} \subseteq S^{(d)}_{all;-x_{d,1}}(j_d)$ such that $P(\tilde{S}^{(d)}) > 0$ (the order of covariates in this Cartesian product coincides with the order of covariates in x).*

Then parameters $\beta_{d,1:L_d}$, $d = 1, \ldots, D$, corresponding to the exclusive covariates in each process are identified.

Condition (b) is a normalization restriction, since in such models parameter vectors can only be identified up to scale and the coefficients $\beta_{d,1}$, $d = 1, \ldots,$ can be normalized to any non-zero values. These normalizations can be different across d. Condition (c), intuitively, requires that for $d = 1, \ldots, D$, there is some some continuous variation in at least one exclusive covariate in x_d, conditional on other covariates, when the other covariates take values from a set of positive measure. We require this condition in the set of x that deliver probabilities $P\left(Y^{(d)} \leq y^{(d)}_{j_d} | x\right)$ strictly between 0 and 1 for some j_d.

Because of the presence of exclusive covariates, we obtain the analogous feature to lattice models, namely that under conditions on Theorem 3, we can identify $\beta_{d,1:L_d}$, the parameters on exclusive covariates, regardless of whether $\beta_{\ell,1:L_\ell}$, $\ell \neq d$, are identified.

Our next result in Theorem 4 strengthens conditions on covariates to obtain the identification of full parameter vectors β_d, $d = 1, \ldots, D$. The identification of parameters corresponding to covariates that are common to at least two indices relies on at least one exclusive covariate in the respective process to have a large support. Large support assumptions are common in the semiparametric literature and , in particular, in semiparametric univariate ordered response models (see e.g. Manski (1985, 1988); Horowitz (2010); Lewbel (2000, 2003)).

Theorem 4 *Suppose all the conditions of Theorem 3 hold. In addition, suppose that if $L_d < k_d$ (that is, there are non-exclusive covariates in x_d), then in condition (c) in Theorem 3:*

(i) *$\underline{x}_{d,1}$ is sufficiently small if the support of ε_d is bounded from above, and*[14]

(ii)

$$\underline{x}_{d,1} = -\infty \tag{8}$$

[14] If $\beta_{d,1}$ was normalized to a negative value, we would require $\overline{x}_{d,1}$ to be sufficiently large if the support of ε_d was bounded from below.

if the support of ε_d *is unbounded from above.*[15]

Then β_d, $d = 1, \ldots, D$, *are identified.*

Our penultimate result concerns the identification of the c.d.f. F.

Theorem 5 *Suppose all the conditions of Theorem 3 hold and in condition (c) in Theorem 3, condition (8) holds for any* $d = 1, \ldots, D$, *and, moreover,*[16]

$$\overline{x}_{d,1} = +\infty. \tag{9}$$

Then under the following normalization for each marginal c.d.f. F_d*:*

$$F_d(e_{0d}) = c_{0d}, \quad d = 1, \ldots, D,$$

for some known e_{0d} *in the support of* ε_d *and some known* $c_{0d} \in (0,1)$, $d = 1, \ldots, D$, *the joint c.d.f.* F *is identified.*

Note that conditions in Theorems 3 - 5 are increasingly more restrictive. For example, in Theorem 5 we require condition (8) for any $d = 1, \ldots, D$, whereas in Theorem 4 we require condition (8) only for d with $L_d < k_d$. Resultantly, coefficients corresponding to exclusive covariates are easier to identify than those corresponding to non-exclusive ones, and that the joint c.d.f. F is harder to identify than index coefficients β_d.

Our final result is on the identification of threshold parameters. This result allows us to find out whether decision-making is consistent with broad bracketing or narrow bracketing. The identification comes from variation in covariates and consideration of probabilities of various rectangular regions such as

$$\begin{aligned} P & \left((Y^{c_1}, \ldots, Y^{c_D}) = \left(y^{(1)}_{j_1}, \ldots, y^{(D)}_{j_D} \right) \mid x \right) = P\left((Y^{*c_1}, \ldots, Y^{*c_D}) \in R_{j_1,\ldots,j_D} | x \right) \\ = & \sum_{\substack{m_1 \in \{0,1\} \ldots \\ m_D \in \{0,1\}}} (-1)^{m_1 + \ldots + m_D} P\left(\bigcap_{d=1}^{D} \left(\varepsilon_d < \alpha^{(d)}_{j_1 - m_1, \ldots, j_{d-1} - m_{d-1}, j_d - m_d, j_{d+1} - m_{d+1}, \ldots, j_D - m_D} - x_d \beta_d \right) \right) \end{aligned}$$

[15] If $\beta_{d,1}$ was normalized to a negative value, we would require $\overline{x}_{d,1} = +\infty$ instead of (8) if the support of ε_d was unbounded from below.

[16] If the support of ε_d is bounded from above, then condition (8) can be replaced with the condition of $x_{d,1}$ taking small enough values, and if the support of of ε_d is bounded from below, then condition (9) can be replaced with the condition of $x_{d,1}$ taking large enough values.

Theorem 6 gives a formal identification result for the thresholds.

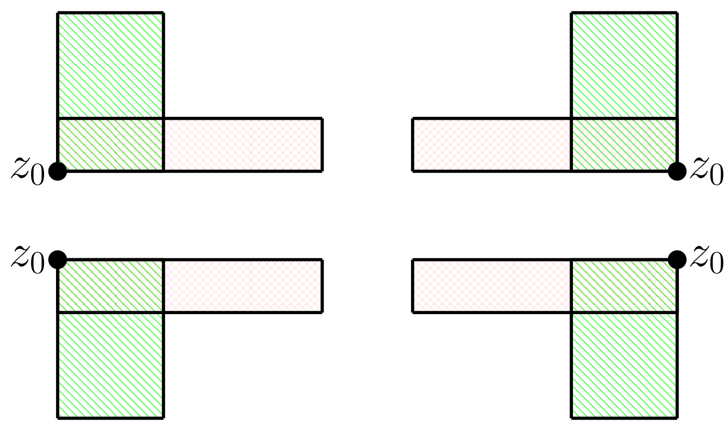
Theorem 6 *Suppose all the conditions of Theorem 5 (including the normalizations for marginal c.d.f.s) hold. Then all the thresholds* $\alpha^{(d)}_{j_1,\cdots,j_{d-1},j_d,j_{d+1},\cdots,j_D}$ *are identified.*

We prove the identification of thresholds in Theorem 6 sequentially. Let us give intuition for the case $D = 2$. First, we identify those thresholds that define regions R_{j_1,j_2} with *both* indices j_d taking an extreme value in that sense that $j_d \in \{0, M_d\}$, $d = 1, 2$. Second, we build on this result to identify thresholds that define regions R_{j_1,j_2} with *only one* index among j_1, j_2 taking an extreme value. At each step in the second stage we proceed sequentially in a way to only having to identify two unknown threshold parameters at a time (rather than three thresholds). We then proceed in an analogous sequential way to cases of rectangular regions "in the middle" having, once again, only two unknown thresholds at a time (rather than three of four). With two known and two unknown thresholds defining a rectangular region R_{j_1,j_2}, the idea behind identification is to hypothetically suppose that there are two pairs of unknown thresholds consistent with observables. We can then think about from which "corner" (corresponds to four combinations of extreme value indices) we want to start the sequential identification process. This would imply that in one of the four situations depicted in Figure 8 the probability masses of the red and green rectangles are exactly the same for any corner point z_0 in the two-dimensional space. The identification comes from the fact that no matter what the support of unobservable $(\varepsilon_1, \varepsilon_2)$ is, at least one of these four depicted situation will be inconsistent with observable probabilities of choice. Thus, there is always a "corner" from which we can start to establish the identification of thresholds. We fill in the details of this identification idea in Appendix B.

6 Parametric assumptions on the distribution of errors

In discrete response models, practitioners often make parametric assumptions on the distribution of unobservables and maintain independence between unobservables and covariates. Once the distributional family of unobservables is specified, sufficient identification conditions are less stringent than those in the semiparametric case. The exact identification conditions usually depend on the parametric family under consideration, and may require some normalizations to eliminate obvious non-identifiability issues.

FIGURE 8: Illustration of the identification idea.



Notes: in each picture, the red and green rectangles have the same probability mass when conjecturing that two different sets of thresholds can generate observables and when proceeding from one of the four "corners".

A popular choice for the distribution of unobservables is the Gaussian, particularly due to ease of modeling correlation among ε_d for $d = 1, \ldots, D$. We formulate this parametric choice in Assumption 5, which combines the normality of unobservables with their independence from the covariates:

Assumption 5 *$(\varepsilon_1, \ldots, \varepsilon_D)$ is independent of $(x_1, \ldots, x_D)$ and has a joint normal distribution $\mathcal{N}(0, \Sigma)$ with*

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1D} \\ \rho_{12} & 1 & \ldots & \rho_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1D} & \rho_{2D} & \ldots & 1 \end{pmatrix}.$$

Assumption 5 already normalizes the means and standard deviations of all ε_d to $\mu_d = 0$ and $\sigma_D = 1$ respectively. This is because identification must use decision probabilities $P\left(Y^{(1)} = y^{(1)}_{j_1}, \ldots, Y^{(D)} = y^{(D)}_{j_D} \mid x\right)$, and if we write down the form of these probabilities using the normality assumption, it follows that

$$(\mu_1, \ldots, \mu_D, \{\alpha^{(d)}_{j_1, j_2, \ldots, j_D}\}, \beta_1, \ldots, \beta_D, \sigma_1, \ldots, \sigma_D, \{\rho_{k_1,k_2}\}_{k_1<k_2})$$

and

$$(\mu_1 + C_1, \ldots, \mu_D + C_D, \{\tilde{C}_d \alpha^{(d)}_{j_1, j_2, \ldots, j_D} + C_d\}, \tilde{C}_1 \beta_1, \ldots, \tilde{C}_D \beta_D, \tilde{C}_1 \sigma_1, \ldots, \tilde{C}_D \sigma_D, \{\rho_{k_1,k_2}\}_{k_1<k_2}),$$

for any $C_d \neq 0$ and $\tilde{C}_d > 0$, $d = 1, \ldots, D$, are observationally equivalent.

In a general non-lattice model satisfying Assumption 5, given sufficient variation in covariates, all parameters of the model (thresholds, index parameters, and correlations) are identified. In a univariate ordered probit model, the index parameter and the thresholds are identified even from sufficient discrete variation in covariates. Undoubtedly, identifying a multivariate non-lattice ordered probit model is more demanding. Appendix C illustrates how to obtain identification in some non-lattice bivariate models. The results rely on an exclusive covariate in at least one latent process. Still, we expect many cases where all parameters are identified even without any exclusive covariates, and we provide an example of that in simulation design 1 in section 8. In that design, the two latent equations contain just one common regressor. Theoretically, clear-cut identification conditions for multivariate non-lattice ordered probit model are difficult to derive. For similar reasons, in multinomial probit models, which consider many latent processes with correlated unobservables, formal identification conditions are not available in the discrete choice literature.

7 Estimation

In this section, we discuss estimation of lattice and non-lattice multivariate ordered response models.

7.1 Estimation in semiparametric models

In what follows, we briefly outline some possibilities for estimating parameters in semiparametric models. A theme of this section is that while existing univariate ordered response estimation methods generalize to lattice models, there are immediate complications in their extension to non-lattice models.

7.1.1 Extending Coppejans (2007)

Coppejans (2007) offers one of the many estimation methods for univariate ordered response models under independence of the error and covariates. In what follows, we extend it to multivariate ordered response models, describing the bivariate case for illustrational simplicity. Suppose we have a random sample $\left\{(y^{(1)(i)}, y^{(2)(i)}, x_1^{(i)}, x_2^{(i)})\right\}_{i=1}^{N}$. The idea is to maximize the log-likelihood function

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j_1=1}^{M_1}\sum_{j_2=1}^{M_2} 1\left[(y^{(1)(i)}, y^{(2)(i)}) = (y_{j_1}^{(1)}, y_{j_2}^{(2)})\right]\log(\ell_{j_1,j_2}^{(i)}), \quad \text{where}$$

$$\begin{aligned} \ell_{j_1,j_2}^{(i)} &= F\left(a_{j_1,j_2}^{(1)} - x_1^{(i)}b_1, a_{j_1 j_2}^{(2)} - x_2^{(i)}b_2\right) - F\left(a_{j_1-1,j_2}^{(1)} - x_1^{(i)}b_1, a_{j_1,j_2}^{(2)} - x_2^{(i)}b_2\right) \\ &- F\left(a_{j_1,j_2}^{(1)} - x_1^{(i)}b_1, a_{j_1,j_2-1}^{(2)} - x_2^{(i)}b_2\right) + F\left(a_{j_1-1,j_2}^{(1)} - x_1^{(i)}b_1, a_{j_1,j_2-1}^{(2)} - x_2^{(i)}b_2\right), \end{aligned}$$

for joint c.d.f. of unobservables F. Coppejans (2007) uses a quadratic B-spline to estimate the c.d.f of unobservables. The multivariate analogy is tensor-product B-splines. For instance, in the bivariate case the tensor-product basis consists of $S_1 \cdot S_2$ products of polynomials $\mathcal{R}$ in the form

$$\mathcal{R}_{1;s_1,S_1}(e_1;q_1)\mathcal{R}_{2;s_2,S_2}(e_2;q_2), \quad s_1 = 1,\ldots,S_1,\ s_2 = 1,\ldots,S_2,$$

here calculated for specific values of e_1 and e_2, with q_d denoting the degree of B-spline in dimension $d = 1,2$. A general tensor-product B-spline, which approximates $F(e_1, e_2)$, is a linear combination of these base tensor-product polynomials with coefficients $\{h_{s_1 s_2}\}$, $s_d = 1,\ldots,S_d$, $d = 1,2$:

$$\sum_{s_1=1}^{S_1}\sum_{s_2=1}^{S_2} h_{s_1 s_2}\mathcal{R}_{1;s_1,S_1}(e_1;q_1)\mathcal{R}_{2;s_2,S_2}(e_2;q_2).$$

The linear constraints

$$\begin{aligned} h_{s_1 s_2} &\le h_{s_1+1,s_2}, \quad \forall\, s_1 = 1,\ldots,S_1-1, \quad s_2 = 1,\ldots,S_2 \\ h_{s_1 s_2} &\le h_{s_1,s_2+1}, \quad \forall\, s_2 = 1,\ldots,S_2-1, \quad s_1 = 1,\ldots,S_1 \end{aligned}$$

guarantee monotonicity of the tensor-product B-spline in each dimension. Additionally, the linear constraints

$$0 \le h_{s_1,s_2} \le 1, \quad \forall\, s_1, s_2$$

guarantee natural c.d.f. bounds of 0 and 1.[17] And linear equality constraints on $h_{s_1 s_2}$ can impose normalization restrictions on F_d. Coherency requires additional constraints on thresholds. As indicated previously, in the bivariate case these constraints have the form given in equation (5)

[17] For more details on shape constraints in tensor-product B-splines, see Bhattacharya and Komarova (2022).

and these can be included as a penalty in the objective through

$$-\lambda_N(\alpha^{(1)}_{j_1+1,j_2} - \alpha^{(1)}_{j_1,j_2})^2 \cdot (\alpha^{(2)}_{j_1,j_2+1} - \alpha^{(2)}_{j_1,j_2})^2 \tag{10}$$

for a large $\lambda_N > 0$.

7.1.2 Extending alternative approaches

The approach in Klein and Sherman (2002), which analyzes the univariate model, estimates the index parameter in the first stage using kernel density estimates of the conditional probability of choosing below a certain level. In the second stage, the approach estimates threshold parameters using shift restrictions. The estimation method extends to multivariate lattice models. The first stage is potentially implementable in the non-lattice context, but difficulty will arise in extending the shift restrictions. The same is true for the more recent approach in Liu and Yu (2019).

The Lewbel (2000) methodology, which in particular applies to univariate ordered response models, can be extended to multivariate lattice models but is difficult, if not impossible, to extend to non-lattice models. The same applies to Lewbel (2003), which focuses on the estimation of thresholds only.

The approach in Chen and Khan (2003) for univariate ordered response models estimates the index parameter subject to a normalization restriction.[18] This approach extends to multivariate lattice models by considering each dimension individually. However, their approach does not immediately generalize to all parameters of non-lattice models without additional assumptions. This is because the equality

$$P\left(Y^{(1)} = y^{(1)}_{j_1}, Y^{(2)} = y^{(2)}_{j_2} \mid x_1, x_2\right) = P\left(Y^{(1)} = y^{(1)}_1, Y^{(2)} = y^{(2)}_1 \mid \tilde{x}_1, \tilde{x}_2\right)$$

only implies that $\tilde{x}_d\beta_d = x_d\beta_d$ if and only if x_{-d} remains fixed, for $d = 1, 2$. Hence, with reference to the discussion in 5.2, the method in Chen and Khan (2003) only estimates parameters on exclusive regressors.

Combining the ideas of pairwise differences (Honoré and Powell, 2005) and maximum rank correlation (MRC) estimation (Han, 1987) will deliver estimates of the parameters $\beta_{d,1:L_d}$ corre-

[18] Their exposition covers the more general case of heteroskedastic errors and therefore applies to homoskedastic models as well.

sponding to exclusive covariates. Pairwise differencing ensures that when we analyze dimension d, we only look at the cases when non-exclusive covariates $x_{d,L_d+1:k_d}$ in dimension d and all other covariates x_ℓ, $\ell \neq d$ are close. The use of maximum rank correlation follows the result and proof of Theorem 3. The proof of that theorem shows that the distribution of $Y^{(d)}$ conditional on $(x_1, \ldots, x_{d,1:L_d}, x_{d,L_d+1:k_d}, \ldots, x_D)$ first order stochastically dominates the distribution of of $Y^{(d)}$ conditional on $(x_1, \ldots, \tilde{x}_{d,1:L_d}, x_{d,L_d+1:k_d}, \ldots, x_D)$ if and only if $x_{d,1:L_d}\beta_{d,1:L_d} > \tilde{x}_{d,1:L_d}\beta_{d,1:L_d}$. Resultantly, the estimation method maximizes the objective function given by

$$
\begin{aligned}
Q_d(b_{d,1:L_d}) &= \sum_{i=1}^{N}\sum_{j>i} 1\left(Y^{(d)(i)} > Y^{(d)(j)}\right) 1\left(x^{(i)}_{d,1:L_d} b_{d,1:L_d} > x^{(j)}_{d,1:L_d} b_{d,1:L_d}\right) \mathcal{K}_{d;H_d}\left(\Delta^{(i),(j)}\right) \\
\Delta^{(i),(j)} &= \left(x^{(i)}_1 - x^{(j)}_1, \ldots, x^{(i)}_{d-1} - x^{(j)}_{d-1}, x^{(i)}_{d,L_d+1:k_d} - x^{(j)}_{d,L_d+1:k_d}, x^{(i)}_{d+1} - x^{(j)}_{d+1}, \ldots, x^{(i)}_D - x^{(j)}_D\right).
\end{aligned}
$$

The term $\mathcal{K}_{d;H_d}$ extracts with a reasonable weight only those i and j whose observations $(x^{(i)}_1, \ldots, x^{(i)}_{d,L_d+1:k_d}, \ldots, x^{(i)}_D)$ and $(x^{(j)}_1, \ldots, x^{(j)}_{d,L_d+1:k_d}, \ldots, x^{(j)}_D)$ are sufficiently close to each other.[19] More formally, $\mathcal{K}_{d;H_d}(z) = |H_d|^{-1/2}\mathcal{K}_d({H_d}^{-1/2}z)$, with $\mathcal{K}_d$ being a $\sum_{\ell\neq d} k_\ell + k_d - L_d$-variate kernel, and H_d being the symmetric and positive definite bandwidth $(\sum_{\ell\neq d} k_\ell) \times (\sum_{\ell\neq d} k_\ell)$ matrix. The maximization of $Q_d(b_{d,1:L_d})$ with respect to $b_{d,1:L_d}$ consistently estimates all $\beta_{d,1:L_d}$.[20] Instead of the MRC estimator, we could incorporate other estimators used in single-index models.

7.2 Parametric estimation

To fix ideas, we discuss parametric probit estimation in the bivariate case ($D = 2$), noting that extensions to other parametric distributions are straightforward.[21] Thus, we suppose that Assumption 5 holds for $D = 2$ so that $(\varepsilon_1, \varepsilon_2)$ are independent of (x_1, x_2) and

$$
\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right). \tag{11}
$$

To start, we introduce some notation. Let α stack $\alpha^{(d)}_{j_1,j_2}$ for $j_d = 1, \ldots, M_d - 1$ and $d = 1, 2$. Define the full set of parameters to estimate as $\theta = (\beta_1', \beta_2', \alpha', \rho)'$. Finally, set the extreme values

[19] This is the idea of Honoré and Powell (2005)

[20] The consistency property of the MRC estimator follows from the first-order stochastic dominance relationship mentioned above.

[21] The choice of other parametric distributions may require different normalization restrictions.

of the thresholds: $\alpha^{(1)}_{M_1,j_2} = \alpha^{(2)}_{j_1,M_2} = +\infty$ and $\alpha^{(1)}_{0,j_2} = \alpha^{(2)}_{j_1,0} = -\infty$.

Then, given a random sample $\left\{(y^{(1)(i)}, y^{(2)(i)}, x_1^{(i)}, x_2^{(i)})\right\}_{i=1}^{N}$ the unconstrained log-likelihood function is

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j_1=1}^{M_1}\sum_{j_2=1}^{M_2} 1\left[(y^{(1)(i)}, y^{(2)(i)}) = (y^{(1)}_{j_1}, y^{(2)}_{j_2})\right]\log(\ell^{(i)}_{j_1,j_2}) = \frac{1}{N}\sum_{i=1}^{N}\log(\ell^{(i)}),$$

where

$$\begin{aligned}\ell^{(i)}_{j_1,j_2} &= \Phi_2\left(\alpha^{(1)}_{j_1,j_2} - x_1^{(i)}\beta_1, \alpha^{(2)}_{j_1 j_2} - x_2^{(i)}\beta_2; \rho\right) - \Phi_2\left(\alpha^{(1)}_{j_1-1,j_2} - x_1^{(i)}\beta_1, \alpha^{(2)}_{j_1,j_2} - x_2^{(i)}\beta_2; \rho\right)\\ &- \Phi_2\left(\alpha^{(1)}_{j_1,j_2} - x_1^{(i)}\beta_1, \alpha^{(2)}_{j_1,j_2-1} - x_2^{(i)}\beta_2; \rho\right) + \Phi_2\left(\alpha^{(1)}_{j_1-1,j_2-1} - x_1^{(i)}\beta_1, \alpha^{(2)}_{j_1-1,j_2-1} - x_2^{(i)}\beta_2; \rho\right),\end{aligned}$$

$\ell^{(i)}$ is equal to $\ell^{(i)}_{j_1,j_2}$ if and only if $(y^{(1)(i)}, y^{(2)(i)}) = (y^{(1)}_{j_1}, y^{(2)}_{j_2})$, and $\Phi_2(\cdot,\cdot;\rho)$ denotes the standard bivariate normal c.d.f. with correlation parameter ρ.

The constrained maximum likelihood estimator (MLE) $\hat{\theta}$ solves the optimisation problem

$$\max_{\theta}\mathcal{L}(\theta) \quad \text{subject to} \quad r(\theta) = 0$$

where $r(\theta)$ stacks the local hierarchical constraints in (5). These constraints are differentiable and so under the typical MLE regularity conditions (Newey and McFadden, 1994), $\hat{\theta}$ is consistent and satisfies

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V),$$

where $V = BJB'$,

$$\begin{aligned} J &= \mathbb{E}\left[\frac{\partial\log(\ell^{(i)})}{\partial\theta}\frac{\partial\log(\ell^{(i)})}{\partial\theta'}\right],\\ B &= J^{-1} - J^{-1}R'(RJ^{-1}R')^{-1}RJ^{-1},\end{aligned}$$

and $R = \dfrac{\partial r(\theta_0)}{\partial\theta'}$. The natural plug-in sample-analogue estimators of J and R provide consistent estimators for the variance-covariance matrix.

Computationally, we found advantages in incorporating the constraints through the penalty term as given in (10). Further, iterative estimation procedures which alternate between estimating thresholds α and separately (β, ρ) via concentrated likelihood may aid in cases with a particularly complex non-lattice structure.

8 Monte Carlo experiments

Now we turn to Monte Carlo simulations. We focus on the parametric case with normal errors as described in section 6 and compare the performance of the newly proposed estimator in 7.2 to the standard bivariate ordered probit estimator (that is, the estimator of the lattice model under normal errors). The baseline model across all simulations is

$$\begin{aligned} Y^{*c_1} &= x\beta_1 + w_1\gamma_1 + \varepsilon_1 \\ Y^{*c_2} &= x\beta_2 + w_2\gamma_2 + \varepsilon_2, \end{aligned}$$

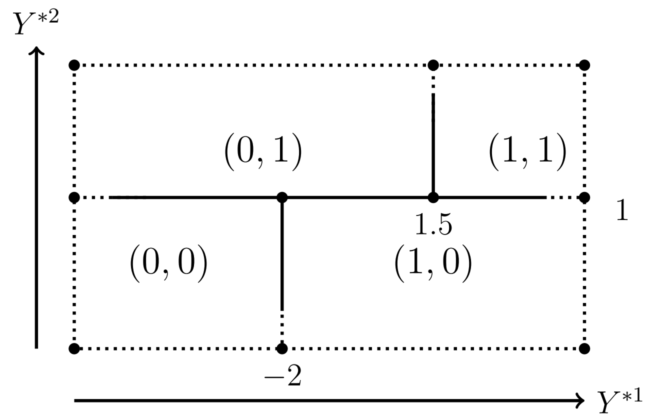
and the unobservables are independent of regressors and jointly normal, as given in (11). This form of a baseline model allows us to differentiate between exclusive and non-exclusive covariates explicitly. We aim to illustrate estimation in different scenarios, such as those with no exclusive covariates ($\gamma_1 = \gamma_2 = 0$) and those with an exclusive covariate in just one latent process. There are two main takeaways from this section. First, in the given parametric model, parameters of a non-lattice model may be identified even without exclusive covariates. Second, estimating lattice models instead of non-lattice ones may ignore the broad bracketing nature of the decision process and result in inconsistent estimators of *all* model parameters. Our simulations and empirical applications show settings in which, for example, we expect a positive correlation between unobservables but estimating a lattice structure delivers a statistically significant negative correlation. The degree of inconsistency in β parameters depends on how well a lattice model approximates the true non-lattice one. In some situations, therefore, the degree of inconsistency of such parameters is mild, whereas in other cases, it is more severe.

For each simulation design, we estimate the parameters on 250 independent random samples. The sample size is 5000 and the penalty term in (10) is set equal to N. Unreported results from alternative choices of N and λ are quantitatively similar. We discuss two designs in text and include a third in appendix D.

Design 1: 2×2 structure, no excluded regressors

We start with a simulation investigating whether we can identify parameters in non-lattice probit models even without exclusion restrictions. To explore this, we set $\gamma_1 = \gamma_2 = 0$ thus effectively removing w_1 and w_2 from the latent equations. We draw the only common regressor x from

FIGURE 9: Latent variable space in design 1



a uniform $[-5, 5]$ distribution. We set $\beta_1 = \beta_2 = 1$, $\rho = 0.33$, and create a 2×2 non-lattice structure with thresholds $\alpha_{01}^{(2)} = \alpha_{11}^{(2)} = 1$ along with $\alpha_{10}^{(1)} = -2$ and $\alpha_{11}^{(1)} = 1.5$. We show the structure in Figure 9.

Table 1 lists the across-simulation means and standard deviations of all parameter estimates. Table 1 shows that the new non-lattice estimation method estimates all parameters with minimal bias. In this case, as we intuitively expect, the bivariate lattice ordered probit method is able to estimate β_1 and the threshold in the first dimension reasonably well, but performs poorly in estimating β_2, ρ, and the threshold in the second dimension. The estimates for ρ are less precise relative to designs 2 and 3 because of the lack of excluded regressors.

Table 1: Simulation results design 1

Parameter	Truth	Non-lattice model	Lattice model
β_1	1	1.00 (0.03)	0.77 (0.019)
β_2	0.5	0.50 (0.02)	0.00 (0.01)
ρ	0.33	0.33 (0.12)	-0.93 (0.02)
$\alpha_{11}^{(2)}$	1	1.00 (0.04) 1.00 (0.04)	0.72 (0.04)
$\alpha_{10}^{(1)}$	-2	-1.99 (0.07)	-0.42 (0.02)
$\alpha_{11}^{(1)}$	1.5	1.50 (0.08)	

Notes: Table 1 reports the sample mean and sample standard deviations (in parentheses) of the estimates of the design 1 parameters, over 250 repeated samples. The "Nonlattice model" column provide estimates from using the newly proposed nonlattice bivariate ordered probit model. The "Lattice model" column assumes a lattice structure, but estimates the two equations jointly.

Design 2: 4×3 with one excluded covariate

In the second simulation design, we extend the number of discrete values M_d in both dimensions. The discrete dependent variable Y^{c_1} can take four values and Y^{c_2} can take three values. This generates a 4×3 non-lattice structure, illustrated in Figure 25. The common covariate x follows a uniform $[-3, 3]$ distribution. The covariate w_1 is a discrete random variable taking values -2.5, -1.5, -0.5 and 0.5 with equal probability 0.25. We set $\gamma_2 = 0$ thus effectively removing w_2 in the second equation.

The parameter values are $\beta_1 = 1.5, \gamma_1 = -4, \beta_2 = 3$ and $\rho = 0.5$. Table 2 lists the across-simulation means and standard deviations of the index parameters and the correlation coefficient. Table 3 in appendix D provides the values of the thresholds, together with their estimated means and standard deviations. The newly proposed method estimates all the parameters with almost no bias. On the contrary, the bivariate lattice ordered probit method estimates the parameters with relatively large bias. The mean squared error of the newly proposed method is far lower than the lattice bivariate probit method for all of the parameters. Assuming a lattice structure makes estimating the correlation parameter ρ decidedly difficult, with the method failing to estimate the correct sign for ρ, let alone an approximately close value.

Table 2: Simulation results design 2

Parameter	Truth	Non-lattice model	Lattice model
β_1	1.5	1.50 (0.04)	0.61 (0.01)
γ_1	-4	-4.01 (0.09)	-2.51 (0.04)
β_2	3	2.99 (0.10)	1.64 (0.03)
ρ	0.5	0.50 (0.06)	-0.60 (0.03)

Notes: Table 2 reports the sample mean and sample standard deviations (in parentheses) of the estimates of the model parameters, over 250 repeated samples. See the notes in table 1 for further details about the columns.

Additional simulation designs are in Appendix D.

9 Applications

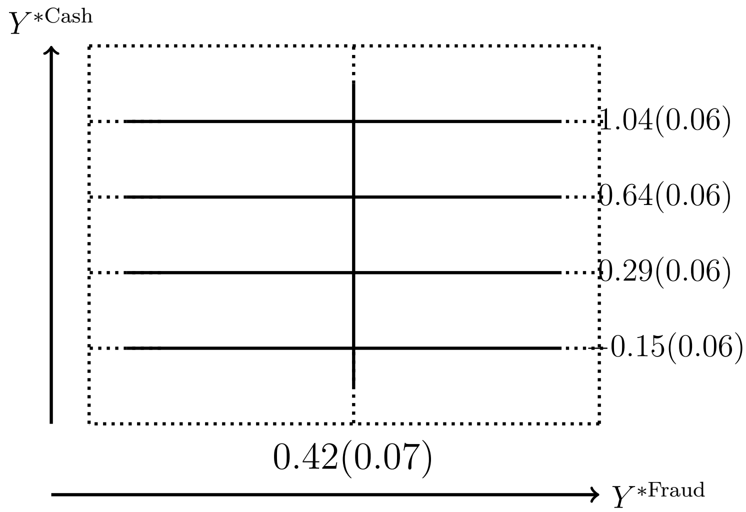
We finish the paper with empirical examples of non-lattice models. We use data from the Survey of Consumer Payment Choice (SCPC) (Foster, Greene, and Stavins, 2021).[22] The Federal Reserve Banks of Atlanta, Boston and San Francisco run the SCPC every October. It is designed primarily to elicit information on American citizens' adoption of various payment instruments. For example, it has recently focused on the vast increase in online and mobile payment methods relative to cash and check payments resulting from the COVID-19 pandemic.

We collect a sample of just over 4,600 surveyed individuals between 2015 and 2020. For these individuals, the survey contains information on their demographics (income, age, gender, and education) and their adoption of various payment choice methods such as credit cards, cryptocurrency and online/mobile payment devices such as Google Pay and PayPal. Individuals report their opinions on the safety, convenience, and cost of various payment methods such as cash, checks, credit and debit cards, and prepaid cards. Finally, individuals report their exposure to fraud, their FICO score bucket, and their role in organizing finances for their household. Foster, Greene, and Stavins (2021) provide further details on the most recent wave of the survey.

We focus on two models: the first investigates the relationship between exposure to identity theft

[22] Other research has used the survey, including Benetton and Compiani (2022) and Kahn and Linares Zegarra (2016)

Figure 10: Estimates from the identity theft example, assuming a lattice model



and opinions on the security of various payment instruments; the second studies the joint decision structures of familiarity with cryptocurrencies and opinion on the value of bitcoin. Appendix D contains another empirical application which studies broad versus narrow bracketing in online payment instrument choice.

We adopt parametric approaches for all three models, assuming that error terms are jointly standard normal with unknown correlation to be estimated.

9.1 Identity theft and opinions on cash

First, we illustrate the use of non-lattice models without reference to bracketing. For Y^1, we use a dummy variable equal to one if the respondent knows anyone—themselves included—who was a victim of identity theft. For Y^2, we use an ordered variable representing the individual's opinion on the security of using cash as a payment method.[23] This variable can take five values ranging from 1 to 5. The answer 1 corresponds to the opinion that cash is a very risky payment instrument, and 5 to the belief that cash is very secure. The covariate vectors x_1 and x_2 in both processes are identical, containing demographics including dummies for low household income, low education status, (non)male gender and a continuous variable representing age.[24]

[23]See Kahn and Linares Zegarra (2016) for a detailed analysis of the relationship between identity theft and payment methods assuming lattice models.

[24]More specifically, low income is 1 if annual household income falls below \$50,000, and low education is 1 if the individual did not attend college.

Figure 11: Estimates from the identity theft example, assuming a non-lattice model

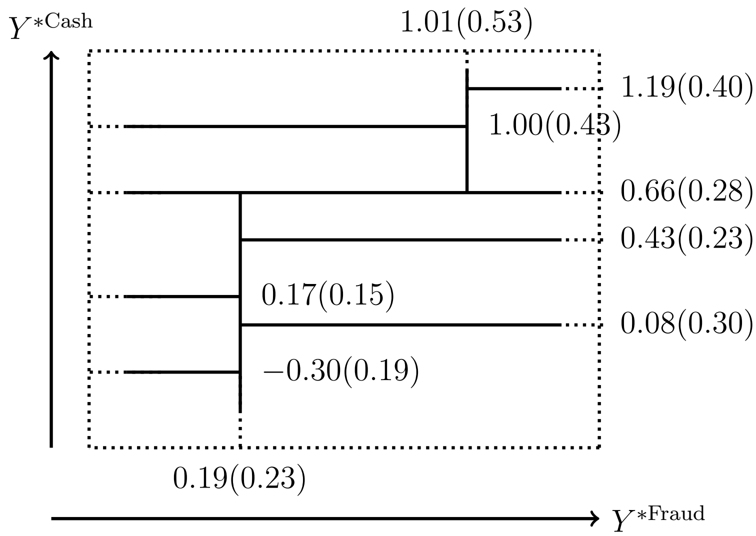
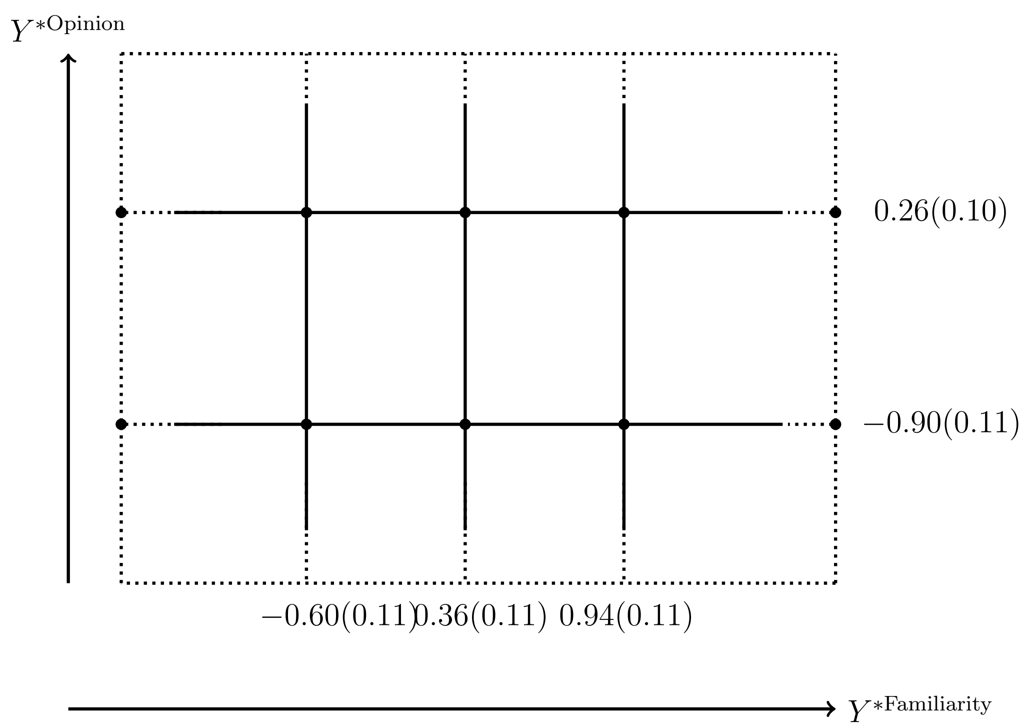


Table 6 in Appendix D presents estimates of β and ρ. The lattice model implies that lower-income and lower-education individuals are more likely to witness identity theft than the non-lattice model. Otherwise, the β coefficients are similar across lattice and non-lattice models. The thresholds in the non-lattice model, as shown in Figure 11, imply that individuals who have been a victim of identity theft have higher thresholds for low values of the cash security variable. This means that victims of identity theft are more likely to think that cash is a safe payment instrument relative to other options such as credit or debit cards. The thresholds also imply that individuals who have a strong opinion in favor of cash have a higher threshold to be a victim of identity theft. Lattice models impose that the thresholds that shape beliefs on cash as a payment method do not have any relationship with individuals' previous exposure to identity theft. In the absence of this relationship, it is unsurprising that the lattice model estimates a negative correlation coefficient (-0.04) compared to the positive coefficient (0.48) estimated in the non-lattice model.

9.2 Cryptocurrency Familiarity and Optimism

In the second application we consider whether the decision structure for opinions on the future value of bitcoin is interdependent with the decision structure on the level of familiarity with cryptocurrencies. In particular, we run the same style of models as above, except with Y^1 as an ordered variable representing *familiarity* with bitcoin. This variable can take four values: -1 if the individual is not at all familiar with bitcoin, 0 if the individual is slightly familiar, 1 if the individual is somewhat familiar and 2 if the individual is either moderately or extremely familiar.[25] For Y^2, we use an ordered variable that reflects opinion on the value of bitcoin in one year. The variable takes the values -1 if the individual believes bitcoin will *decrease* in value, 0 if the individual believes it will stay the same and 1 if the individual believes its value will *increase*. Table 7 in Appendix D presents the estimates of β and ρ. The value of ρ ranges from 0.03 in the lattice model to 0.84 in the non-lattice model. The coefficients are also markedly different, with most coefficients differing more than 20% in relative magnitude and the coefficient on male differing in sign. The lattice model estimates a negative coefficient on male in the opinion equation. A negative coefficient means that, conditional on other covariates, males are more likely to believe that the value of bitcoin will decrease relative to non-males, implying pessimism.[26] The non-lattice model reverses this, finding what many expect to be more likely: that males are more bullish than non-males on the future value of bitcoin.

FIGURE 12: Estimates from cryptocurrency example, assuming a lattice model



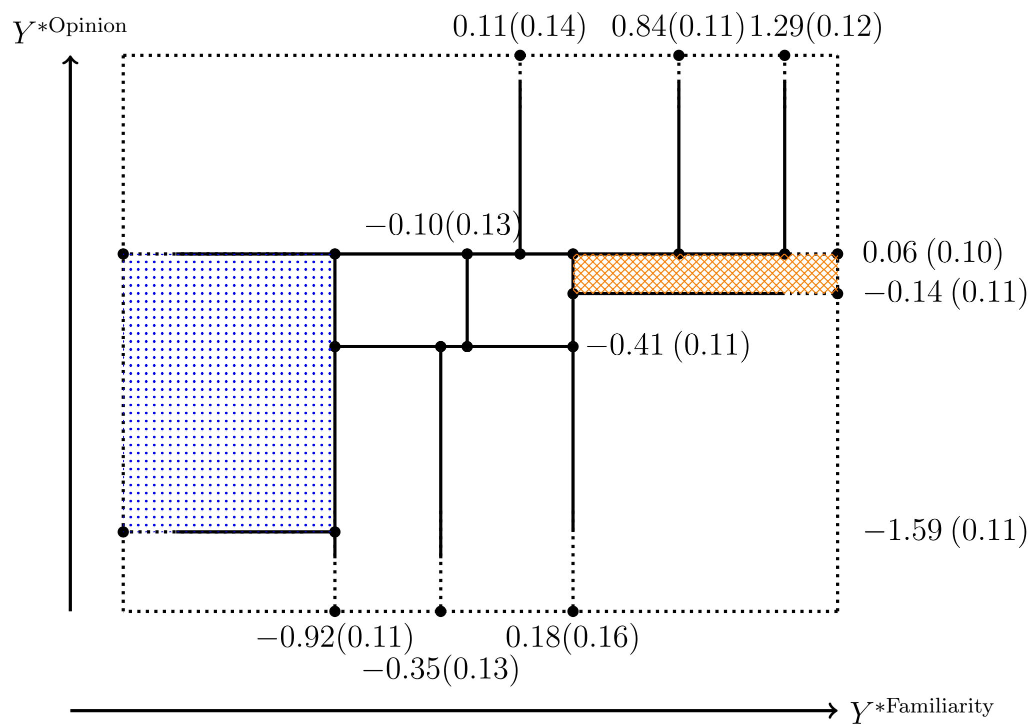
Finally, our interest turns to the estimation of the thresholds, in particular in the non-lattice model. Figures 12 and 13 illustrate the estimated thresholds in lattice and non-lattice models respectively. The lattice model forces that the thresholds—which determine opinion on the future

[25]We combine moderate and extremely familiar because only a few individuals reported that they were extremely familiar with bitcoin.

[26]This coefficient is, admittedly, rather imprecisely estimated.

value of bitcoin—do not change with the individual's familiarity with bitcoin. The non-lattice model, not forced to employ this rigid structure, finds a wide range (blue dots) in the lattice space representing "no change in value" for individuals with little familiarity with bitcoin. This result seems reasonable, and in fact, one would expect the default position of someone with little understanding of bitcoin to be that its value would remain roughly the same. On the other hand, for those with particular familiarity with bitcoin, there is a small region (red crosshatch) in the lattice space representing "no change in value". Individuals with a familiarity of bitcoin are more likely to take a stance on the future value of bitcoin, be it a positive or a negative one.

FIGURE 13: Estimates from cryptocurrency example when assuming a non-lattice model.



10 Conclusion

This paper introduces a general model of multivariate ordered discrete response, with a focus on lattice, non-lattice, and hierarchical classes. We give formal identification results in the semiparametric case and offer estimation approaches for semiparametric and parametric formulations. Several extensions warrant investigation. For example, future work can relax the homoskedasticity assumption or consider median independence of unobservables, which may suit a partial identification approach. We also encourage an extensive analysis of the generalizability of existing univariate lattice semiparamteric methods to non-lattice models. Finally, there are opportunities for empirical applications of non-lattice models in cases where lattice models are inappropriate and, perhaps more interestingly, in other settings where the degree of broad/narrow bracketing is not obvious *prima facie*.

References

The numbers at the end of every reference link to the pages citing the reference.

Abbring, J. H. and J. J. Heckman (2007): "Chapter 72: Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," *Handbook of Econometrics*, 6, 5145–5303.

Agresti, A. (2010): *Analysis of Ordinal Categorical Data*, John Wiley & Sons.

Aradillas-López, A. and A. M. Rosen (2022): "Inference in ordered response games with complete information," *Journal of Econometrics*, 226, 451–476.

Benetton, M. and G. Compiani (2022): "Investors' Beliefs and Cryptocurrency Prices," *Working Paper.*

Berry, S. and P. Reiss (2007): "Empirical Models of Entry and Market Structure," *Handbook of Industrial Organization*, 3, 1845–1886.

Besley, T. and T. Persson (2011): "The Logic of Political Violence," *The Quarterly Journal of Economics*, 126, 1411–1445.

Bhat, C. R. and V. Pulugurta (1998): "A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions," *Transportation Research Part B: Methodological*, 32, 61–75.

Bhattacharya, D. and T. Komarova (2022): "Incorporating social welfare in program-evaluation and treatment choice," *Working paper.*

Boes, S. and R. Winkelmann (2006): "Ordered Response Models," in *Modern Econometric Analysis*, Springer, chap. 12, 167–181.

Buliung, R. N. and P. S. Kanaroglou (2007): "Activity–Travel Behaviour Research: Conceptual Issues, State of the Art, and Emerging Perspectives on Behavioural Analysis and Simulation Modelling," *Transport Reviews*, 27, 151–187.

Camara, M. (2021): "Computationally Tractable Choice," *Working Paper*.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997): "Labor Supply of New York City Cabdrivers: One Day at a Time," *The Quarterly Journal of Economics*, 112, 407–441.

Cameron, S. V. and J. J. Heckman (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106, 262–333.

Carneiro, P., K. Hansen, and J. Heckman (2003): "Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice," *International Economic Review*, 44.

Chen, S. and S. Khan (2003): "Rates of convergence for estimating regression coefficients in heteroskedastic discrete response models," *Journal of Econometrics*, 117, 245–278.

Chesher, A. and A. M. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85, 959–989.

——— (2020): "Structural modeling of simultaneous discrete choice," *Cemmap Working Paper*.

Chiappori, P. and B. Salanie (2000): "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 56–78.

Cicchetti, C. J. and J. A. Dubin (1994): "A Microeconometric Analysis of Risk Aversion and the Decision to Self-Insure," *Journal of Political Economy*, 102, 169–186.

Ciliberto, F. and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

Cohen, A. (2005): "Asymmetric Information and Learning: Evidence from the Automobile Insurance Market," *The Review of Economics and Statistics*, 87, 197–207.

Coppejans, M. (2007): "On efficient estimation of the ordered response model," *Journal of Econometrics*, 137, 577–614.

Cunha, F., J. J. Heckman, and S. Navarro (2007): "The Identification and Economic Content of Ordered Choice Models with Stochastic Thresholds," *International Economic Review*, 48, 1273–1309.

Ellis, A. and D. J. Freeman (2020): "Revealing Choice Bracketing," *Working Paper*.

Fang, H., M. P. Keane, and D. Silverman (2008): "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market," *Journal of Political Economy*, 116, 303–350.

Filer, R. and M. Honig (2005): "Endogenous Pensions and Retirement Behaviour," *Working Paper*.

Finkelstein, A. and K. McGarry (2006): "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market," *American Economic Review*, 96.

Finkelstein, A. and J. Poterba (2004): "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market," *Journal of Political Economy*, 112, 183–208.

Foster, K., C. Greene, and J. Stavins (2021): "The 2020 Survey of Consumer Payment Choice: Summary Results," *Federal Reserve Bank of Atlanta Research Data Reports*.

Genius, M., C. J. Pantzios, and V. Tzouvelekas (2006): "Information Acquisition and Adoption of Organic Farming Practices," *Journal of Agricultural and Resource Economics*, 31, 93–113.

Greene, W. H. and D. A. Hensher (2010): *Modeling Ordered Choices: A Primer*, Cambridge University Press.

Han, A. K. (1987): "Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator," *Journal of Econometrics*, 35, 303–316.

Heckman, J. J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–959.

Heckman, J. J., R. J. Lalonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," *Handbook of Labor Economics*, 3, Part A, 1865–2097.

Heckman, J. J. and S. Navarro (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136, 341–396.

HERRNSTEIN, R. J. AND D. PRELEC (1991): "Melioration: A Theory of Distributed Choice," *The Journal of Economic Perspectives*, 5, 137–156.

HEYMAN, G. M. (1996): "Resolving the contradictions of addiction," *Behavioral and Brain Sciences*, 19, 561–574.

HONORE, B. E. AND A. DE PAULA (2010): "Interdependent Durations," *The Review of Economic Studies*, 77, 1138–1163.

HONORÉ, B. E. AND J. L. POWELL (2005): *Pairwise Difference Estimators for Nonlinear Models*, Cambridge University Press, 520–553.

HOROWITZ, J. L. (2010): *Semiparametric and Nonparametric Methods in Econometrics*, Springer Series in Statistics, Springer-Verlag.

KAHN, C. AND J. M. LINARES ZEGARRA (2016): "Identity Theft and Consumer Payment Choice: Does Security Really Matter?" *Journal of Financial Services Research*, 50, 121–159.

KAHNEMAN, D. AND D. LOVALLO (1993): "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking," *Management Science*, 39, 17–31.

KLEIN, R. W. AND R. P. SHERMAN (2002): "Shift Restrictions and Semiparametric Estimation in Ordered Response Models," *Econometrica*, 70, 663–691.

LEE, M.-J. (1992): "Median regression for ordered discrete response," *Journal of Econometrics*, 51, 59–77.

LEWBEL, A. (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–6177.

——— (2003): "Ordered Response Threshold Estimation," *Unpublished Working Paper*.

LIAN, C. (2020): "A Theory of Narrow Thinking," *The Review of Economic Studies*, 88, 2344–2374.

LIU, R. AND Z. YU (2019): "Simple Semiparametric Estimation of Ordered Response Models: with an Application to the Interdependence Duration Models," *Tsukuba Economics Working Papers*.

MALMENDIER, U. AND S. NAGEL (2011): "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *The Quarterly Journal of Economics*, 126, 373–416.

Manski, C. (1988): "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83, 729–738.

Manski, C. F. (1975): "The Maximum Score Estimator of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205–228.

——— (1985): "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27, 313–333.

Newey, W. K. and D. McFadden (1994): "Chapter 36: Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 4, 2111–2245.

Rabin, M. and G. Weizsäcker (2009): "Narrow Bracketing and Dominated Choices," *American Economic Review*, 99, 1508–43.

Read, D., G. Loewenstein, and M. Rabin (1999): "Choice Bracketing," *Journal of Risk and Uncertainty*, 19, 171–197.

Scott, D. M. and P. S. Kanaroglou (2002): "An activity-episode generation model that captures interactions between household heads: development and empirical analysis," *Transportation Research Part B: Methodological*, 36, 875–896.

Simonson, I. and R. S. Winer (1992): "The Influence of Purchase Quantity and Display Format on Consumer Preference for Variety," *Journal of Consumer Research*, 19, 133–138.

Small, K. (1987): "A Discrete Choice Model for Ordered Alternatives," *Econometrica*, 55, 409–24.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165.

Thakral, N. and L. T. Tô (2021): "Daily Labor Supply and Adaptive Reference Points," *American Economic Review*, 111, 2417–43.

Thaler, R. H. (1999): "Mental accounting matters," *Journal of Behavioral Decision Making*, 12, 183–206.

Tversky, A. and D. Kahneman (1981): "The Framing of Decisions and the Psychology of Choice," *Science*, 211, 453–458.

Wang, X. and S. Chen (2022): "Partial Identification and Estimation of Semiparametric Ordered Response Models with Interval Regressor Data," *Oxford Bulletin of Economics and Statistics*, 84, 830–849.

Zhang, M. (2021): "A Theory of Choice Bracketing under Risk," *Working Paper.*

A Coherency

Even though coherency is usually only a concern in strategic interaction settings, which we do not consider, we discuss it because coherency is employs in our identification strategy and, for a more practical perspective, coherency constraints need to be imposed in estimation. Throughout this discussion, we consider a general non-lattice model with discrete responses $\{(y^{(1)}_{j_1},\dots,y^{(D)}_{j_D})\}$, $j_d = 1,\dots,M_d$, $d=1,\dots,D$, a set of thresholds $\alpha^{(d)}_{j_1,\dots,j_{d-1},j_d,j_{d+1},\dots,j_D}$, $j_d=1,\dots,M_d$, $d=1,\dots,D$, and normalization constraints (2) and (3).

By *coherency of the model with the given set of thresholds in the latent space* we mean that rectangles $R_{j_1,\dots,j_D}$ defined in (1) form a partition of $\mathbb{R}^D$ – that is, they are mutually exclusive and give the whole $\mathbb{R}^D$ in the union. By *coherency of the model in the observable probabilities* we means that the sum of probabilities of all discrete responses $(y^{(1)}_{j_1},\dots,y^{(D)}_{j_D})$ conditional on observables is equal to one.

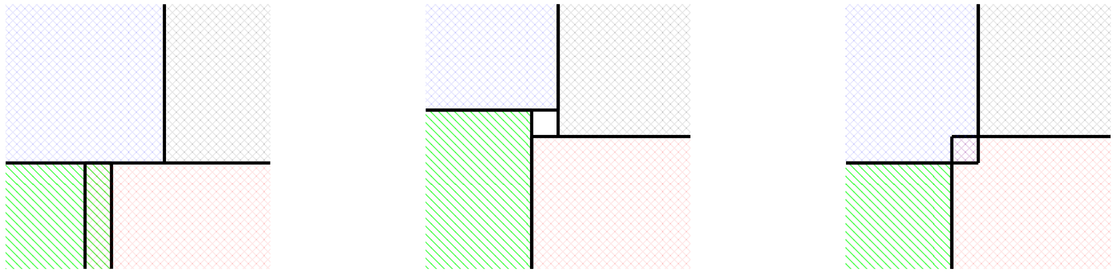
The question is under what conditions the coherency of the model in observable probabilities translates into the coherency in the latent space. We are interested in *generic coherency* in the latent space. If $\mathcal{O}$ denotes observables in the model, then by the model being generically coherent in the latent space we mean that the coherency has to hold for *any set of* $\prod_{d=1}^{D} M_d$ *observable choice probabilities* $P\left(\cap_{d=1}^{D}(Y^{c_d}=y^{(d)}_{j_d})\,|\,\mathcal{O}\right)$, $j_d=1,\dots,M_d$, $d=1,\dots,D$, *which are non-negative and sum up to one.*

We consider the cases of $D=2$ and $D=3$ in turn.

A.1 Bivariate case

Theorem 7 *If $D=2$, a non-lattice model with given discrete responses and a given set of thresholds is generically coherent in the latent space if and only if it is locally hierarchical.*

FIGURE 14: Potential violations of coherency.



Panel 1 Panel 2 Panel 3

Notes: Violation of coherency in Panel 1 is ruled out by the thresholds structure in (1). Violations of coherency in Panels 2 and 3 are not immediately ruled out by (1).

Proof of Theorem 7. Before we proceed to proving sufficiency and necessity directions, we note that the structure of our model captured in the definition of rectangles R_{j_1,j_2} already imposes certain restrictions on the ways coherency in the latent space may be violated. For example, the coherency violation in Panel 1 in Figure 14 is eliminated by the definition of the rectangles R_{j_1,j_2} in (1) but the coherency violations in Panels 2 and 3 are not.

Sufficiency. Suppose the model is locally hierarchical. This means that each local decision model is hierarchical. The recursive definition of hierarchical models implies that in the local decision model, conditional on a subset in $\mathbb{R}^2$, we are dealing with a partitioning of that subset into 2^2 regions. The local model given in Definition 4 and which only looks at 2^2 adjacent responses $(y^{(1)}_{j_1+\ell_1}, y^{(2)}_{j_2+\ell_2})$, $\ell_d \in \{0,1\}$, $d = 1,2$, is *the only local model* that involves thresholds $\alpha^{(d)}_{j_1,j_2}$ for both $d = 1,2$, for this particular j_1, j_2. Considering all local models, at every step we will find coherency associated with rectangles surrounding each node $\left(\alpha^{(1)}_{j_1,j_2}, \alpha^{(2)}_{j_1,j_2}\right)$. From this we are able to make a conclusion about the coherency of the whole model.

Necessity. Now consider a coherent model. Suppose it is not locally hierarchical. This means that there is a local decision model which is not hierarchical, and this holds for any realization of observed choice probabilities $P\left(\cap_{d=1}^2 (Y^{c_d} = y^{(d)}_{j_d}) \,|\, \mathcal{O}\right)$, $j_d = 1, \ldots, M_d$, $d = 1,2$, which are non-negative and sum up to 1.

We can focus on the 2×2 local decision model of choosing among responses $(y^{(1)}_{j_1+\ell_1}, y^{(2)}_{j_2+\ell_2})$, $\ell_d \in \{0,1\}$, $d = 1,2$, and the coherency condition for this local problem can be written explicitly in the form

$$(\alpha^{(1)}_{j_1,j_2} - \alpha^{(1)}_{j_1,j_2+1})(\alpha^{(2)}_{j_1,j_2} - \alpha^{(2)}_{j_1+1,j_2}) = 0.$$

Suppose this condition is violated. Suppose, for example, that $\alpha^{(1)}_{j_1,j_2} < \alpha^{(1)}_{j_1,j_2+1}$, $\alpha^{(2)}_{j_1,j_2} < \alpha^{(2)}_{j_1+1,j_2}$. Then the rectangles for regions (j_1, j_2+1) and (j_1+1, j_2) have an intersection with a non-empty interior $(\alpha^{(1)}_{j_1,j_2+1} - \alpha^{(1)}_{j_1,j_2}) \times (\alpha^{(2)}_{j_1+1,j_2} - \alpha^{(2)}_{j_1,j_2}) \subset \mathbb{R}^2$, which violates coherency when all observed conditional decision probabilities are strictly positive. If $\alpha^{(1)}_{j_1,j_2} > \alpha^{(1)}_{j_1,j_2+1}$, $\alpha^{(2)}_{j_1,j_2} > \alpha^{(2)}_{j_1+1,j_2}$, then the rectangles for regions (j_1, j_2) and (j_1+1, j_2+1) have an intersection with a non-empty interior $(\alpha^{(1)}_{j_1,j_2} - \alpha^{(1)}_{j_1,j_2+1}) \times (\alpha^{(2)}_{j_1,j_2} - \alpha^{(2)}_{j_1+1,j_2}) \subset \mathbb{R}^2$, which once again violates coherency when all observed conditional decision probabilities are strictly positive. If $\alpha^{(1)}_{j_1,j_2} < \alpha^{(1)}_{j_1,j_2+1}$, $\alpha^{(2)}_{j_1,j_2} > \alpha^{(2)}_{j_1+1,j_2}$, then the region $(\alpha^{(1)}_{j_1,j_2+1} - \alpha^{(1)}_{j_1,j_2}) \times (\alpha^{(2)}_{j_1,j_2} - \alpha^{(2)}_{j_1+1,j_2})$ does not belong to any decision rectangles $R_{j_1+\ell_1,j_2+\ell_2}$, $\ell_1, \ell_2 \in \{0,1\}$. This contradicts coherency when the distribution of $(Y^{*c_1}, Y^{*c_2}) \,|\, \mathcal{O}$ has $\mathbb{R}^2$ as the support, implying that the sum of all conditional probabilities of choice has to be strictly less than 1. The case of $\alpha^{(1)}_{j_1,j_2} > \alpha^{(1)}_{j_1,j_2+1}$, $\alpha^{(2)}_{j_1,j_2} < \alpha^{(2)}_{j_1+1,j_2}$ is analogous to the last case just described. $\square$

The sufficiency proof in Theorem 7 can be easily suitably extended to any $D \geq 2$ thus giving the result that a model being locally hierarchical implies coherency.

A.2 Trivariate case

Next, we consider the case $D = 3$. In this case the property of being locally hierarchical is a sufficient condition but not a necessary one, as will be evident from the discussion below. Resultantly, coherent decision making has a more complicated characterization to that in case $D = 2$.

To see what other situations beyond being locally hierarchical are possible in $D = 3$, consider a local model deciding among responses $(y^{(1)}_{j_1+\ell_1}, y^{(2)}_{j_2+\ell_2}, y^{(3)}_{j_3+\ell_3})$, $\ell_d \in \{0,1\}$, $d = 1,2,3$. Suppose this local model is not hierarchical. If the first stage in the definition of the hierarchical model for this local model goes through but the second or a later stage breaks down it means we are in effectively the situation when local hierarchical is violated for $D = 2$ and we have already obtained a contradiction for that case. Thus, it makes sense to consider only cases when the definition of hierarchical breaks down at the first step. In other words, for dimension $d = 3$ thresholds $\alpha^{(3)}_{j_1+\ell_1,j_2+\ell_2,j_3}$ are not invariant to $\ell_1, \ell_2 \in \{0,1\}$, and analogous conclusions can be made for $d = 1,2$. Continue to take $d = 3$. Among four thresholds $\alpha^{(3)}_{j_1+\ell_1,j_2+\ell_2,j_3}$, $\ell_d \in \{0,1\}$, there is a highest one.

<u>Situation 1</u>. Consider a situation when there are exactly two highest among these four thresholds.

Let these two highest thresholds correspond to two adjacent orthants – without a loss of generality R_{j_1+1,j_2,j_3+1} and R_{j_1+1,j_2+1,j_3+1} (corresponding to decisions $(y^{(1)}_{j_1+1}, y^{(2)}_{j_2}, y^{(3)}_{j_3+1})$ and $(y^{(1)}_{j_1+1}, y^{(2)}_{j_2+1}, y^{(3)}_{j_3+1})$, respectively).

a) If in dimension 1 thresholds $\alpha^{(1)}_{j_1,j_2,j_3+1}$ and $\alpha^{(1)}_{j_1,j_2+1,j_3+1}$ are the same, then to ensure coherency of the local model thresholds in dimension 1 have to remain the same when we "move down" in dimension 3 and consider orthants R_{j_1+1,j_2,j_3} and R_{j_1+1,j_2+1,j_3} corresponding to decisions $(y^{(1)}_{j_1+1}, y^{(2)}_{j_2}, y^{(3)}_{j_3})$ and $(y^{(1)}_{j_1+1}, y^{(2)}_{j_2+1}, y^{(3)}_{j_3})$, respectively (easy to indicate a contradiction otherwise). This gives the independence of the threshold $\alpha^{(1)}_{j_1,j_2+\ell_2,j_3+\ell_3}$ from (ℓ_2, ℓ_3) implying that for our local model the first step in the definition of the hierarchical model goes through. This contradicts our supposition.

b) If $\alpha^{(1)}_{j_1,j_2,j_3+1}$ and $\alpha^{(1)}_{j_1,j_2+1,j_3+1}$ are different, then to ensure coherency in dimensions 1 and 2 when the response in the third dimension remains fixed as $y^{(3)}_{j_3+1}$, it has to hold that $\alpha^{(2)}_{j_1,j_2,j_3+1} = \alpha^{(2)}_{j_1+1,j_2,j_3+1}$. But then the fact that there are two highest thresholds implies that $\alpha^{(2)}_{j_1,j_2,j_3} = \alpha^{(2)}_{j_1+1,j_2,j_3} = \alpha^{(2)}_{j_1,j_2,j_3+1} = \alpha^{(2)}_{j_1+1,j_2,j_3+1}$ – that is, the thresholds in dimension 2 remain the same when we "move down" in dimension 3 and consider orthants R_{j_1+1,j_2,j_3} and R_{j_1+1,j_2+1,j_3}. This gives the independence of the threshold $\alpha^{(2)}_{j_1+\ell_1,j_2,j_3+\ell_3}$ from j_1 and j_3 implying that for our local model the first step in the definition of the hierarchical model goes through. This contradicts our supposition.

If these two highest thresholds correspond to two non-adjacent orthants, then (analogously to a) and b) above) the condition that ensures coherency in dimensions 1 and 2 when the response in dimension 3 is fixed as $y^{(3)}_{j_3+1}$ has to continue to hold when the response in dimension 3 is fixed as $y^{(3)}_{j_3}$. This will lead to us showing that the first step in the recursive definition of the hierarchical model for our local model goes through.

Situation 2. Next, consider a situation when the highest threshold among $\alpha^{(3)}_{j_1+\ell_1,j_2+\ell_2,j_3}$, $\ell_d \in \{0, 1\}$, is strictly greater than the other three. Without a loss of generality suppose it is associated with $\ell_1 = 0$, $\ell_2 = 0$ (in other words, it is associated with responses $y^{(1)}_{j_1}$ and $y^{(2)}_{j_2}$ in the first two dimensions). Then the coherency of the model (when we "move up" from the orthant R_{j_1,j_2,j_3} to R_{j_1,j_2,j_3+1}) implies that

$$\alpha^{(1)}_{j_1,j_2,j_3} = \alpha^{(1)}_{j_1,j_2,j_3+1}, \quad \alpha^{(2)}_{j_1,j_2,j_3} = \alpha^{(2)}_{j_1,j_2,j_3+1}. \tag{12}$$

Fixing response $y^{(3)}_{j_3}$ in the third dimension, from the case $D = 2$ we conclude that due to coherency in the projection on that case,

$$(\alpha^{(1)}_{j_1,j_2+1,j_3} - \alpha^{(1)}_{j_1,j_2,j_3})(\alpha^{(2)}_{j_1+1,j_2,j_3} - \alpha^{(2)}_{j_1,j_2,j_3}) = 0. \tag{13}$$

Fixing response $y^{(3)}_{j_3+1}$ in the third dimension, from the case $D = 2$ we conclude that due to coherency in the projection on that case,

$$(\alpha^{(1)}_{j_1,j_2+1,j_3+1} - \alpha^{(1)}_{j_1,j_2,j_3+1})(\alpha^{(2)}_{j_1+1,j_2,j_3+1} - \alpha^{(2)}_{j_1,j_2,j_3+1}) = 0. \tag{14}$$

First, consider a subcase when (13) and (14) are guaranteed by equalities of thresholds in the same dimensions – for concreteness, suppose that

$$\alpha^{(1)}_{j_1,j_2+1,j_3} - \alpha^{(1)}_{j_1,j_2,j_3} = 0, \quad \alpha^{(1)}_{j_1,j_2+1,j_3+1} - \alpha^{(1)}_{j_1,j_2,j_3+1} = 0,$$

then also taking into account (12), one can easily conclude that

$$\alpha^{(1)}_{j_1,j_2+1,j_3+1} = \alpha^{(1)}_{j_1,j_2+1,j_3} = \alpha^{(1)}_{j_1,j_2,j_3+1} = \alpha^{(1)}_{j_1,j_2,j_3},$$

which means that the first step in the recursive definition of the hierarchical model can indeed be applied to our local model by considering dimension 1. This contradicts our earlier supposition about the recursive definition breaking down at the first step.

Second, consider a subcase when (13) and (14) are guaranteed by equalities of thresholds in different dimensions – for concreteness, suppose that

$$\alpha^{(1)}_{j_1,j_2+1,j_3} - \alpha^{(1)}_{j_1,j_2,j_3} = 0, \quad \alpha^{(1)}_{j_1,j_2+1,j_3+1} - \alpha^{(1)}_{j_1,j_2,j_3+1} \neq 0,$$

$$\alpha^{(2)}_{j_1+1,j_2,j_3} - \alpha^{(2)}_{j_1,j_2,j_3} \neq 0, \quad \alpha^{(2)}_{j_1+1,j_2,j_3+1} - \alpha^{(2)}_{j_1,j_2,j_3+1} = 0.$$

Also taking into account (12) we have

$$\alpha^{(1)}_{j_1,j_2,j_3} = \alpha^{(1)}_{j_1,j_2+1,j_3} = \alpha^{(1)}_{j_1,j_2,j_3+1} \neq \alpha^{(1)}_{j_1,j_2+1,j_3+1}, \quad \alpha^{(2)}_{j_1,j_2,j_3} = \alpha^{(2)}_{j_1+1,j_2,j_3+1} = \alpha^{(2)}_{j_1,j_2,j_3+1} \neq \alpha^{(2)}_{j_1+1,j_2,j_3}.$$

Thus, the characterizations we obtain here describe coherent decision structures which are not locally hierarchical. There are several potential ways to interpret them. One interpretation is that when taking a local problem and any two dimensions our of three, say dimensions d_1 and d_2, there is a response $(y^{(d_1)}_{j_{d_1}+\ell_{d_1}}, y^{(d_2)}_{j_{d_2}+\ell_{d_2}})$ such that the binary decision of choosing that response in dimensions d_1 and d_2 is independent of the realization of the latent utility in the third dimension d_3. Another, perhaps more technical interpretation is that when taking a local problem and focusing on any dimensions (d_1, d_2) there are only two different two-dimensional non-lattice structures whenever we take the latent utility value in the third dimension as given – this is something violated in the local hierarchical models.

Situation 3. A situation when there are three highest thresholds among $\alpha^{(3)}_{j_1+\ell_1,j_2+\ell_2,j_3}$, $\ell_d \in \{0,1\}$, is isomorphic to Situation 2 and we can consider it analogously by considering one lowest threshold among those.

Thus, as we can see, coherent structures in dimensions $D \geq 3$ strictly nest local hierarchical structures, thus, permitting richer decision making structures. A more practical question is how to characterize them through mathematical relations involving thresholds that can be directly taken to an estimation procedure. To characterize coherency, we once again focus on each local problem of choosing among 2^3 discrete responses $(y^{(1)}_{j_1+\ell_1}, y^{(2)}_{j_2+\ell_2}, y^{(3)}_{j_3+\ell_3})$, $\ell_d \in \{0,1\}$, $d = 1,2,3$. Each choice corresponds to a 3-dimensional rectangle which has three non-infinite sides and three infinite ones (due to locality, we disregard other choices). Naturally, we can refer to such rectangles as orthants. Among these 2^3 orthants we consider 2^2 pairs of "opposite" orthants. By "opposite" orthants we understand orthants corresponding to discrete response $(y^{(1)}_{j_1+\ell_1}, y^{(2)}_{j_2+\ell_2}, y^{(3)}_{j_3+\ell_3})$ and $(y^{(1)}_{j_1+1-\ell_1}, y^{(2)}_{j_2+1-\ell_2}, y^{(3)}_{j_3+1-\ell_3})$ for given $(\ell_1,\ell_2,\ell_3) \in \{0,1\}^3$. For concreteness, take "opposite" orthants corresponding to $(y^{(1)}_{j_1}, y^{(2)}_{j_2}, y^{(3)}_{j_3})$ and $(y^{(1)}_{j_1+1}, y^{(2)}_{j_2+1}, y^{(3)}_{j_3+1})$. There orthants are, respectively, $\times_{d=1}^{3}(-\infty, \alpha^{(d)}_{j_1,j_2,j_3}]$ and $(\alpha^{(1)}_{j_1,j_2+1,j_3+1}, +\infty) \times (\alpha^{(2)}_{j_1+1,j_2,j_3+1}, +\infty) \times (\alpha^{(3)}_{j_1+1,j_2+1,j_3}, +\infty)$. A part of coherency requirement is that in at least one dimension the non-infinite thresholds for these two orthants have to be the same – that is,

$$\left(\alpha^{(1)}_{j_1,j_2,j_3} - \alpha^{(1)}_{j_1,j_2+1,j_3+1}\right) \cdot \left(\alpha^{(2)}_{j_1,j_2,j_3} - \alpha^{(2)}_{j_1+1,j_2,j_3+1}\right) \cdot \left(\alpha^{(3)}_{j_1,j_2,j_3} - \alpha^{(3)}_{j_1+1,j_2+1,j_3}\right) = 0. \tag{15}$$

An analogous requirement would apply to the other three pairs of "opposite" orthants. Thus, additionally we will require

$$\left(\alpha^{(1)}_{j_1,j_2,j_3+1} - \alpha^{(1)}_{j_1,j_2+1,j_3}\right) \cdot \left(\alpha^{(2)}_{j_1,j_2,j_3+1} - \alpha^{(2)}_{j_1+1,j_2,j_3}\right) \cdot \left(\alpha^{(3)}_{j_1,j_2,j_3} - \alpha^{(3)}_{j_1+1,j_2+1,j_3}\right) = 0 \tag{16}$$

(when considering orthants for $(y_{j_1}^{(1)}, y_{j_2}^{(2)}, y_{j_3+1}^{(3)})$ and $(y_{j_1+1}^{(1)}, y_{j_2+1}^{(2)}, y_{j_3}^{(3)})$);

$$\left(\alpha_{j_1,j_2+1,j_3}^{(1)} - \alpha_{j_1,j_2,j_3+1}^{(1)}\right) \cdot \left(\alpha_{j_1,j_2,j_3}^{(2)} - \alpha_{j_1+1,j_2,j_3+1}^{(2)}\right) \cdot \left(\alpha_{j_1,j_2+1,j_3}^{(3)} - \alpha_{j_1+1,j_2,j_3}^{(3)}\right) = 0 \tag{17}$$

(when considering orthants for $(y_{j_1}^{(1)}, y_{j_2+1}^{(2)}, y_{j_3}^{(3)})$ and $(y_{j_1+1}^{(1)}, y_{j_2}^{(2)}, y_{j_3+1}^{(3)})$); and

$$\left(\alpha_{j_1,j_2,j_3}^{(1)} - \alpha_{j_1,j_2+1,j_3+1}^{(1)}\right) \cdot \left(\alpha_{j_1+1,j_2,j_3}^{(2)} - \alpha_{j_1,j_2,j_3+1}^{(2)}\right) \cdot \left(\alpha_{j_1+1,j_2,j_3}^{(3)} - \alpha_{j_1,j_2+1,j_3}^{(3)}\right) = 0 \tag{18}$$

(when considering orthants for $(y_{j_1+1}^{(1)}, y_{j_2}^{(2)}, y_{j_3}^{(3)})$ and $(y_{j_1}^{(1)}, y_{j_2+1}^{(2)}, y_{j_3+1}^{(3)})$). In addition to constraints (15)-(18), we also impose analogous constraints for two-dimensional rectangles in each of the two dimensions when the discrete response in the third dimension is fixed. These are, of course, well familiar coherency constraints from the case $D = 2$:

$$\begin{aligned} \left(\alpha_{j_1,j_2,j_3}^{(1)} - \alpha_{j_1,j_2+1,j_3}^{(1)}\right) \cdot \left(\alpha_{j_1+1,j_2,j_3}^{(2)} - \alpha_{j_1,j_2,j_3}^{(2)}\right) &= 0, \\ \left(\alpha_{j_1,j_2,j_3+1}^{(1)} - \alpha_{j_1,j_2+1,j_3+1}^{(1)}\right) \cdot \left(\alpha_{j_1+1,j_2,j_3+1}^{(2)} - \alpha_{j_1,j_2,j_3+1}^{(2)}\right) &= 0 \end{aligned} \tag{19}$$

(in dimensions 1 and 2);

$$\begin{aligned} \left(\alpha_{j_1,j_2,j_3}^{(3)} - \alpha_{j_1,j_2,j_3+1}^{(3)}\right) \cdot \left(\alpha_{j_1+1,j_2,j_3}^{(1)} - \alpha_{j_1,j_2,j_3}^{(1)}\right) &= 0, \\ \left(\alpha_{j_1,j_2+1,j_3}^{(3)} - \alpha_{j_1,j_2+1,j_3+1}^{(3)}\right) \cdot \left(\alpha_{j_1+1,j_2+1,j_3}^{(1)} - \alpha_{j_1,j_2+1,j_3}^{(1)}\right) &= 0 \end{aligned} \tag{20}$$

(in dimensions 1 and 3); and

$$\begin{aligned} \left(\alpha_{j_1,j_2,j_3}^{(3)} - \alpha_{j_1,j_2,j_3+1}^{(3)}\right) \cdot \left(\alpha_{j_1,j_2,j_3}^{(2)} - \alpha_{j_1,j_2+1,j_3}^{(2)}\right) &= 0, \\ \left(\alpha_{j_1+1,j_2,j_3}^{(3)} - \alpha_{j_1+1,j_2,j_3+1}^{(3)}\right) \cdot \left(\alpha_{j_1+1,j_2,j_3}^{(2)} - \alpha_{j_1+1,j_2+1,j_3}^{(2)}\right) &= 0 \end{aligned} \tag{21}$$

(in dimensions 2 and 3). Thus, for each local problem we have to impose constraints (15)-(21).

Since these constraints when implemented for each local problem are necessary and sufficient for coherency, they of course incorporate the local hierarchical case as well as the other coherent cases discussed above.

Cases $D > 3$ are considered analogously to $D = 3$.

B Proofs

We start by introducing some additional notations we will use in some of the proofs.

Additional notations. The survival function of ε is denoted as $\overline{F}$:

$$\overline{F}(z_1, \dots, z_D) = P\left(\cap_{d=1}^{D}(\varepsilon_d > z_d)\right).$$

For the two components of ε – say, ε_d and ε_h, notations $F_{\overline{d},h}$, $F_{d,\overline{h}}$ and $F_{\overline{d},\overline{h}}$ are defined as follows:

$$\begin{aligned}
F_{\overline{d},h}(z_d, z_h) &= P\left(\varepsilon_d > z_d, \varepsilon_h \le z_h\right),\\
F_{d,\overline{h}}(z_d, z_h) &= P\left(\varepsilon_d \le z_d, \varepsilon_h > z_h\right),\\
F_{\overline{d},\overline{h}}(z_d, z_h) &= P\left(\varepsilon_d > z_d, \varepsilon_h > z_h\right).
\end{aligned}$$

Analogously, for any subvector $(\varepsilon_{d_1}, \dots, \varepsilon_{d_S})$, $F_{d_1,\dots,d_S}$ will dente the c.d.f. of this subvector. When some indices among subscripts appear with the bar (as $\overline{d}_s$), this will mean that the event for the respective ε_{d_s} is the "survival" event $\{\varepsilon_{d_s} > z_{d_s}\}$. Thus, for instance,

$$\begin{aligned}
F_{\overline{1},2,3}(z_1, z_2, z_3) &= P\left(\varepsilon_1 > z_1, \varepsilon_2 \le z_2, \varepsilon_3 \le z_3\right),\\
F_{\overline{1},\overline{2},3}(z_1, z_2, z_3) &= P\left(\varepsilon_1 > z_1, \varepsilon_2 > z_2, \varepsilon_3 \le z_3\right),\ \textit{etc.}
\end{aligned}$$

B.1 Proof of Theorem 1

Fix a dimension d, $d = 1, \dots, D$, for which the condition of this theorem holds. Because of the lattice structure and Assumption 1 we have for any $x_d \in \mathcal{X}_d$,

$$P(Y^{c_d} \le y_j^{(d)} \,|\, x_d) = F_d\left(\alpha_j^{(d)} - x_d\beta_d\right), \quad j = 1, \dots, M_d. \tag{22}$$

Assumption 3 guarantees that $P(Y^{c_d} \le y_j^{(d)}|x_d)$ will not be degenerate for $x_d \in S_d$ (in the sense that it will not take values 0 or 1 only). Relation (22) is the basis of the identification strategy. Strict monotonicity of c.d.f. F_d automatically gives us that for two $x_d.\tilde{x}_d \in S^{(d)}$,

$$P(Y^{c_d} \le y_j^{(d)} \,|\, \tilde{x}_d) > P(Y^{c_d} \le y_j^{(d)} \,|\, x_d) \text{ for some } j \quad \iff \quad \tilde{x}_d\beta_d < x_d\beta_d.$$

Thus, the identification is similar to the one in single-index models with a monotone link function (e.g.

see Manski (1988) for the statistical independence case or Manski (1985) (Lemma 2) for the proof under large support). Notice that we do not need a large support condition for this result.

Take $b_d \neq \beta_d$ (both are normalized in the same way so $b_{d,m(d)} = 1$ and $\beta_{d,m(d)} = 1$). The condition of the theorem implies that there exists a positive measure of $x^0_{d,-m(d)} \in S^{(d)}_{-m(d)}$ such that $x^0_{d,-m(d)}\beta_{-m(d)} \neq x^0_{d,-m(d)}b_{-m(d)}$. Without a loss of generality suppose that $x^0_{d,-m(d)}\beta_{-m(d)} > x^0_{d,-m(d)}b_{-m(d)}$. For any $x^0_{d,m(d)}$ that complements $x^0_{d,-m(d)}$ to a point in $S^{(d)}$ we clearly have $x^0_{d,m(d)} + x^0_{d,-m(d)}\beta_{-m(d)} > x^0_{d,m(d)} + x^0_{d,-m(d)}b_{-m(d)}$. We can take $x^0_{d,m(d)} \in (\underline{x}_{d,m(d)}, \overline{x}_{d,m(d)})$.

Due to the continuity of the regressor $x_{d,m(d)}$ on $(\underline{x}_{d,m(d)}, \overline{x}_{d,m(d)})$, one can find $\tilde{x}^0_{d,m(d)}$ slightly different from $x^0_{d,m(d)}$ such that $(\tilde{x}^0_{d,m(d)}, x^0_{d,-m(d)}) \in S^{(d)}$ and

$$x^0_{d,m(d)} + x^0_{d,-m(d)}\beta_{d,-m(d)} \overset{(*)}{>} \tilde{x}^0_{d,m(d)} + x^0_{d,-m(d)}\beta_{d,-m(d)} \overset{(**)}{>} x^0_{d,m(d)} + x^0_{d,-m(d)}b_{d,-m(d)}$$

If b and β were *both* consistent with the observables, we would have from the inequality (*) that

$$P\left(Y^{c_d} \leq y^{(d)}_j \,|\, (x^0_{d,m(d)}, x^0_{d,-m(d)})\right) < P\left(Y^{c_d} \leq y^{(d)}_j \,|\, (\tilde{x}^0_{d,m(d)}, x^0_{d,-m(d)})\right), \tag{23}$$

and from inequality (**) that

$$P\left(Y^{c_d} \leq y^{(d)}_j \,|\, (\tilde{x}^0_{d,m(d)}, x^0_{d,-m(d)})\right) < P\left(Y^{c_d} \leq y^{(d)}_j \,|\, (x^0_{d,m(d)}, x^0_{d,-m(d)})\right). \tag{24}$$

Inequalities (23) and (24) give a contradiction for the probability on the left-hand side of (23). This contradiction is obtained for a positive measure of $(x^0_{d,m(d)}, x^0_{d,-m(d)})$). This implies that β_d is identified relative to b_d. $\square$

B.2 Proof of Theorem 2

Fix a dimension d, $d = 1, \ldots, D$, for which the condition of this theorem holds. Also fix $j = 1, \ldots, M_d - 1$. Then because of the large support conditions in the theorem one can find two different values $x \in S^{(d;j)}$ and $\tilde{x} \in S^{(d;j+1)}$ such that

$$F_d\left(\alpha^{(d)}_j - x_d\beta_d\right) = F_d\left(\alpha^{(d)}_{j+1} - \tilde{x}_d\beta_d\right).$$

In terms of observables this can described as finding $x \in S^{(d;j)}$ and $\tilde{x} \in S^{(d;j+1)}$ such that $P(Y^{c_d} \leq y^{(d)}_j \,|\, x_d)$ and $P(Y^{c_d} \leq y^{(d)}_{j+1} \,|\, \tilde{x}_d)$ are strictly between 0 and 1. Using the convexity of the support of ε_d

in Assumption 1 and, thus, strict monotonicity of F_d in the interior, we conclude right away that

$$\alpha_{j+1}^{(d)} - \alpha_j^{(d)} = \tilde{x}_d\beta_d - x_d\beta_d.$$

Since β_d is already identified by Theorem 1, we immediately conclude that $\alpha_{j+1}^{(d)} - \alpha_j^{(d)}$ is identified for any $j = 1, \ldots, M_d - 1$. $\square$

B.3 Proof of Theorem 3.

We start by fixing $d = 1, \ldots, D$, and analyzing the marginal probability $P(Y^{c_d} \le y_j^{(d)} \,|\, x)$ for some $j = 1, \ldots, M_d$. For instance, for $d = 1$ we have

$$\begin{aligned} P(Y^{c_1} \le y_j^{(1)} \,|\, x) &= \sum_{\tilde{j}=1}^{j} \sum_{j_2=1}^{M_2} \cdots \sum_{j_D=1}^{M_D} P\left((Y^{*c_1}, \ldots, Y^{*c_D}) \in R_{\tilde{j}, j_2, \ldots, j_D} | x \right) \\ &= \sum_{\tilde{j}=1}^{j} \sum_{j_2=1}^{M_2} \cdots \sum_{j_D=1}^{M_D} P\left((x_1\beta_1 + \varepsilon_1, \ldots, x_D\beta_D + \varepsilon_D) \in R_{\tilde{j}, j_2, \ldots, j_D} | x \right) \end{aligned}$$

We first prove the following lemma.

Lemma 1 *$P(Y^{c_1} \le y_j^{(1)} \,|\, x)$ is non-increasing in $x_1\beta_1$ when other indices $x_\ell\beta_\ell$, $\ell \neq 1$, remain fixed, $j = 1, \ldots, M_1$.*

Proof of Lemma 1. To gain some intuition for this, consider first the case of $D = 2$. In this case,

$$\begin{aligned} P(Y^{c_1} \le y_1^{(1)} \,|\, x_1, x_2) &= \sum_{j_2=1}^{M_2} P\left((x_1\beta_1 + \varepsilon_1, x_2\beta_2 + \varepsilon_2) \in R_{1,j_2} | x_1.x_2 \right) \\ &= \sum_{j_2=1}^{M_2} P(\varepsilon_1 \le \alpha_{1,j_2}^{(1)} - x_1\beta_1, \alpha_{1,j_2-1}^{(2)} - x_2\beta_2 < \varepsilon_2 \le \alpha_{1,j_2}^{(2)} - x_2\beta_2). \end{aligned}$$

Since $\alpha_{1,j_2}^{(2)} > \alpha_{1,j_2-1}^{(2)}$ and $x_2\beta_2$ is fixed, then each $P(\varepsilon_1 \le \alpha_{1,j_2}^{(1)} - x_1\beta_1, \alpha_{1,j_2-1}^{(2)} - x_2\beta_2 < \varepsilon_2 \le \alpha_{1,j_2}^{(2)} - x_2\beta_2)$ is non-increasing in $x_1\beta_1$. Thus, $P(Y^{c_1} \le y_1^{(1)} \,|\, x_1, x_2)$ is non-increasing in $x_1\beta_1$ as well.

The next probability we want to consider is $P(Y^{c_1} \le y_2^{(1)} \,|\, x_1, x_2)$. Due to the coherency of our model (and, hence, the partitioning structure in the decision rule) and normalization restrictions (3), we have that

$$\bigcup_{\tilde{j}=1}^{2} \bigcup_{j_2=1}^{M_2} R_{\tilde{j}, j_2} = \bigcup_{j_2=1}^{M_2} R^*_{2, j_2},$$

where $R^*_{2,j_2} = \left(-\infty, \alpha^{(1)}_{2,j_2}\right] \times \left(\alpha^{(2)}_{2,j_2-1}, \alpha^{(2)}_{2,j_2}\right]$. This implies that

$$\begin{aligned} P(Y^{c_1} \leq y_2^{(1)} \,|\, x_1, x_2) &= P\left((x_1\beta_1 + \varepsilon_1, x_2\beta_2 + \varepsilon_2) \in \cup_{\tilde{j}=1}^{2} \cup_{j_2=1}^{M_2} R_{\tilde{j},j_2} | x_1.x_2\right) \\ &= \sum_{j_2=1}^{M_2} P(\varepsilon_1 \leq \alpha^{(1)}_{2,j_2} - x_1\beta_1, \alpha^{(2)}_{2,j_2-1} - x_2\beta_2 < \varepsilon_2 \leq \alpha^{(2)}_{2,j_2} - x_2\beta_2). \end{aligned}$$

Since $\alpha^{(2)}_{2,j_2} > \alpha^{(2)}_{2,j_2-1}$ for each $j_2 \geq 1$ and $x_2\beta_2$ is fixed, then each $P(\varepsilon_1 \leq \alpha^{(1)}_{2,j_2} - x_1\beta_1, \alpha^{(2)}_{2,j_2-1} - x_2\beta_2 < \varepsilon_2 \leq \alpha^{(2)}_{2,j_2} - x_2\beta_2)$ is non-increasing in $x_1\beta_1$.

Analogously, we can show that $P(Y^{c_1} \leq y_j^{(1)} \,|\, x_1, x_2)$ is non-increasing in $x_1\beta_1$ for any $j = 1, \ldots, M_1$.

Let us now consider the case of any D. Once again, let us start with $P(Y^{c_1} \leq y_1^{(1)} \,|\, x)$:

$$\begin{aligned} P(Y^{c_1} \leq y_1^{(1)} \,|\, x) &= \sum_{j_2=1}^{M_2} \cdots \sum_{j_D=1}^{M_D} P\left((x_1\beta_1 + \varepsilon_1, x_2\beta_2 + \varepsilon_2, \ldots, x_D\beta_D + \varepsilon_D) \in R_{1,j_2,\ldots,j_D}\right) = \\ &= \sum_{j_2=1}^{M_2} \cdots \sum_{j_D=1}^{M_D} \left(F\left(\alpha^{(1)}_{1,j_2,\ldots,j_D} - x_1\beta_1, \alpha^{(2)}_{1,j_2,\ldots,j_D} - x_2\beta_2, \ldots, \alpha^{(D)}_{1,j_2,\ldots,j_D} - x_D\beta_D\right) + \right. \\ &\qquad \left. + \overline{F}\left(-\infty, \alpha^{(2)}_{1,j_2-1,\ldots,j_D} - x_2\beta_2, \ldots, \alpha^{(D)}_{1,j_2,\ldots,j_D-1} - x_D\beta_D\right) - 1\right) \end{aligned}$$

(we we remind the readers that $\overline{F}$ is the joint survival function of $(\varepsilon_1, \ldots, \varepsilon_D)$), which is clearly non-increasing in $x_1\beta_1$ when other indices $x_\ell\beta_\ell$ remain fixed. For any $j = 1, \ldots, M_1$, the partitioning structure in the decision rule guarantees that

$$\bigcup_{\tilde{j}=1}^{j} \bigcup_{j_2=1}^{M_2} \cdots \bigcup_{j_D=1}^{M_D} R_{\tilde{j},j_2,\ldots,j_D} = \bigcup_{j_2=1}^{M_2} R^*_{j,j_2,\ldots,j_D},$$

where

$$R^*_{j,j_2,\ldots,j_D} = \left(-\infty, \alpha^{(1)}_{j,j_2,\ldots,j_D}\right] \bigtimes_{d=2}^{D} \left(\alpha^{(d)}_{j,j_2,\ldots,j_{d-1},j_d-1,j_{d+1},\ldots,j_D}, \alpha^{(d)}_{j,j_2,\ldots,j_{d-1},j_d,j_{d+1},\ldots,j_D}\right].$$

In turn, this gives

$$\begin{aligned} P(Y^{c_1} \leq y_j^{(1)} \,|\, x) &= \sum_{j_2=1}^{M_2} \cdots \sum_{j_D=1}^{M_D} P\left((x_1\beta_1 + \varepsilon_1, x_2\beta_2 + \varepsilon_2, \ldots, x_D\beta_D + \varepsilon_D) \in R^*_{j,j_2,\ldots,j_D}\right) = \\ &= \sum_{j_2=1}^{M_2} \cdots \sum_{j_D=1}^{M_D} \left(F\left(\alpha^{(1)}_{j,j_2,\ldots,j_D} - x_1\beta_1, \alpha^{(2)}_{j,j_2,\ldots,j_D} - x_2\beta_2, \ldots, \alpha^{(D)}_{j,j_2,\ldots,j_D} - x_D\beta_D\right) + \right. \\ &\qquad \left. + \overline{F}\left(-\infty, \alpha^{(2)}_{1,j_2-1,\ldots,j_D} - x_2\beta_2, \ldots, \alpha^{(D)}_{1,j_2,\ldots,j_D-1} - x_D\beta_D\right) - 1\right), \end{aligned}$$

which is obviously non-increasing in $x_1\beta_1$ when other indices $x_\ell\beta_\ell$, $\ell \geq 2$, remain fixed. $\square$

Now we can rely on the results of Lemma 1 to prove Theorem 3.

For simplicity consider $d = 1$ and choose j_1 such that $S^{(1)}_{all}(j_1)$ satisfies Assumption 4. Let's take $b \in \mathbb{R}^{k_1}$ such that $b_1 = 1$ (normalization given in the Theorem). If $L_1 = 1$, then the result of the theorem is already established. Suppose $L_1 > 1$ and $b_{1,2:L_1} \neq \beta_{1,2:L_1}$. Then

$$x_{1,2:L_1}\beta_{1,2:L_1} \neq x_{1,2:L_1}b_{1,2:L_1}$$

for a positive measure of $x_{1,2:L_1}$ that belong to the projection of $S^{(1)}_{all}(j_1)$ on the last first $2 : L_1$ covariates in vector x_1 (note that here we do not employ $x_{1,1}$). Without a loss of generality, suppose that for this positive measure of $x_{1,2:L_1}$ we have

$$x_{1,2:L_1}\beta_{1,2:L_1} > x_{1,2:L_1}b_{1,2:L_1}. \tag{25}$$

Now fix any $x_{1,2:L_1}$ that satisfies (25). Then for any $\tilde{x}_{1,1} \in (\underline{x}_{1,1}.\overline{x}_{1,1})$, we have

$$\tilde{x}_{1,1} + x_{1,2:L_1}\beta_{1,2:L_1} > \tilde{x}_{1,1} + x_{1,2:L_1}b_{1,2:L_1}.$$

Because of some continuous variation in $x_{1,1}$ on $(\underline{x}_{1,1}.\overline{x}_{1,1})$ we can find $\tilde{\tilde{x}}_{1,1} \in (\underline{x}_{1,1}.\overline{x}_{1,1})$ such that

$$\tilde{x}_{1,1} + x_{1,2:L_1}\beta_{1,2:L_1} \overset{(a)}{>} \tilde{\tilde{x}}_{1,1} + x_{1,2:L_1}\beta_{1,2:L_1} \overset{(b)}{>} \tilde{x}_{1,1} + x_{1,2:L_1}b_{1,2:k_1}. \tag{26}$$

Now fix other components in $(x_{1,L_1+1:k_1}, x_2, \ldots, x_D)$ such that

$$(\tilde{x}_{1,1}, x_{1,2:L_1}, x_{1,L_1+1:k_1}, x_2, \ldots, x_D) \in S^{(1)}_{all}(j_1)$$

for the overall collection of covariates.

Notice that because of $x_{1,1}$ being exclusive for $Y^{*(c_1)}$, when we vary $x_{1,1}$, the values of $x_2, \ldots, x_D$ remain exactly the same. This means that in the expression for $P(Y^{c_1} \leq y^{(1)}_j \,|\, x_1, \ldots, x_D)$ (see Lemma 1), the values of $\alpha^{(2)}_{1,1,\ldots,1} - x_2\beta_2, \ldots, \alpha^{(D)}_{1,1,\ldots,1} - x_D\beta_D$ remain exactly the same. This means that by varying $x_{1,1}$, we can equivalently express the ordering of $P(Y^{c_1} \leq y^{(1)}_j \,|\, x_1, \ldots, x_D)$ with the ordering of the first argument in $x_1\beta_1$, as established in Lemma 1. Therefore, (a) in (26) implies that

$$P\left(Y^{c_1} \leq y^{(1)}_j \mid (\tilde{x}_{1,1}, x_{1,2:L_1}, x_2, \ldots, x_D)\right) < P\left(Y^{c_1} \leq y^{(1)}_j \mid (\tilde{\tilde{x}}_{1,1}, x_{1,2:L_1}, x_2, \ldots, x_D)\right).$$

If we assume that both β and b can generate observable choice probabilities of choice, then (b) in (26) implies that

$$P\left(Y^{c_1} \le y_j^{(1)} \mid (\tilde{\tilde{x}}_{1,1}, x_{1,2:L_1}, x_2, \ldots, x_D)\right) < P\left(Y^{c_1} \le y_j^{(1)} \mid (\tilde{x}_{1,1}, x_{1,2:k_1}, x_2, \ldots, x_D)\right).$$

Combining the last two inequalities results in an obvious contradiction

$$P\left(Y^{c_1} \le y_j^{(1)} \mid (\tilde{x}_{1,1}, x_{1,2:L_1}, x_2, \ldots, x_D)\right) < P\left(Y^{c_1} \le y_j^{(1)} \mid (\tilde{x}_{1,1}, x_{1,2:L_1}, x_2, \ldots, x_D)\right),$$

and from our discussion it is clear that this contradiction is obtained for a positive measure of $(\tilde{x}_{1,1}, x_{1,2:L_1}, x_2, \ldots, x_D)$. Therefore, $\beta_{1,2:L_1}$ is identified relative to any $b_{1,2:L_1} \neq \beta_{1,2:L_1}$.

The identification of $\beta_{d,1:L_d}$ (up to normalization $\beta_{d,1} = 1$) for $d = 2, \ldots, D$, is proven analogously. $\square$

B.4 Proof of Theorem 4.

For example, consider $d = 1$ and analyze the marginal probability $P(Y^{c_1} \le y_{j_1}^{(1)} \,|\, x_1, \ldots, x_D)$ for some j_1 that satisfies Assumption 4. As indicated in the proof of Theorem 3,

$$\begin{aligned} P(Y^{c_1} \le y_{j_1}^{(1)} \,|\, x) &= \sum_{\tilde{j}=1}^{j_1} \sum_{j_2=1}^{M_2} \cdots \sum_{j_D=1}^{M_D} P\left((Y^{*c_1}, \ldots, Y^{*c_D}) \in R_{\tilde{j}, j_2, \ldots, j_D} | x\right) \\ &= \sum_{\tilde{j}=1}^{j} \sum_{j_2=1}^{M_2} \cdots \sum_{j_D=1}^{M_D} P\left((x_1\beta_1 + \varepsilon_1, \ldots, x_D\beta_D + \varepsilon_D) \in R_{\tilde{j}, j_2, \ldots, j_D} | x\right) \end{aligned}$$

Suppose, for simplicity, that $L_p < k_p$ for each $p = 2, \ldots, D$. Then, by the condition of the theorem, we can take $x_{p,1} \to -\infty$ for $p = 2, \ldots, D$. Since variable $x_{p,1}$ is exclusive to process p, the value of the index $x_1\beta_1$ remains the same. If $j_2 = 1, \ldots, j_D = 1$, then

$$P\left((x_1\beta_1 + \varepsilon_1, \ldots, x_D\beta_D + \varepsilon_D) \in R_{\tilde{j}, j_2, \ldots, j_D} | x\right) \to F_1\left(\alpha_{\tilde{j}, 1, \ldots, 1}^{(1)} - x_1\beta_1\right).$$

If $j_p > 1$ for some $p = 2, \ldots, D$, then

$$P\left((x_1\beta_1 + \varepsilon_1, \ldots, x_D\beta_D + \varepsilon_D) \in R_{\tilde{j}, j_2, \ldots, j_D} | x\right) \to 0.$$

Thus,

$$P(Y^{c_1} \le y_{j_1}^{(1)} \,|\, x) \to F_1\left(\alpha_{j_1, 1, \ldots, 1}^{(1)} - x_1\beta_1\right) \quad \text{as} \quad x_{2,1} \to -\infty, \ldots, x_{D,1} \to -\infty.$$

Now, in the limit, we can compare the values of $P(Y^{c_1} \le y^{(1)}_{j_1} \,|\, x)$ for different x_1:

$$\lim_{\substack{x_{2,1}\to-\infty,\dots\\ x_{D,1}\to-\infty}} P(Y^{c_1} \le y^{(1)}_{j_1} \,|\, (\widetilde{x}_1, x_2, \dots, x_D)) > \lim_{\substack{x_{2,1}\to-\infty,\dots\\ x_{D,1}\to-\infty}} P(Y^{c_1} \le y^{(1)}_{j_1} \,|\, (\widetilde{\widetilde{x}}_1, x_2, \dots, x_D))$$
$$\iff \quad \widetilde{x}_1\beta_1 < \widetilde{\widetilde{x}}_1\beta_1. \quad (27)$$

Using the continuity of the first covariate in x_1 and the fact that the coefficient $\beta_{1,1}$ is normalized, we can use the same techniques as in Theorem 3 to show that the system of linear inequalities constructed as in (27) identifies β_1.

If for some $d \neq 1$, the support of ε_d is bounded from above, then the condition "$x_{d,1} \to -\infty$" can be replaced with "$x_{d,1}$ take small enough values", as at small enough values of $x_{d,1}$ we will have that $\alpha^{(d)}_{j_1,\dots,j_d,\dots,j_D} - x_d\beta_d$ is above the upper support point of ε_d.

Now consider the case when for some $p = 2, \dots, D$, we have $L_p = k_p$ and, thus, for such p all the covariates in x_p are exclusive to the pth latent process. For convenience, suppose that $L_2 = k_2$ and $L_p < k_p$, $p = 3, \dots, D$. Then if $j_3 = 1, \dots, j_D = 1$, then

$$\begin{aligned}
P\left((x_1\beta_1 + \varepsilon_1, \dots, x_D\beta_D + \varepsilon_D) \in R_{\widetilde{j}_1,\widetilde{j}_2,1,\dots,j_D} | x\right) \to \\
F_{1,2}\left(\alpha^{(1)}_{\widetilde{j}_1,\widetilde{j}_2,1,\dots,1} - x_1\beta_1, \alpha^{(2)}_{\widetilde{j}_1,\widetilde{j}_2,1,\dots,1} - x_2\beta_2\right) - \\
F_{1,2}\left(\alpha^{(1)}_{\widetilde{j}_1-1,\widetilde{j}_2,1,\dots,1} - x_1\beta_1, \alpha^{(2)}_{\widetilde{j}_1,\widetilde{j}_2,1,\dots,1} - x_2\beta_2\right) - \\
F_{1,2}\left(\alpha^{(1)}_{\widetilde{j}_1,\widetilde{j}_2,1,\dots,1} - x_1\beta_1, \alpha^{(2)}_{\widetilde{j}_1,\widetilde{j}_2-1,1,\dots,1} - x_2\beta_2\right) + \\
F_{1,2}\left(\alpha^{(1)}_{\widetilde{j}_1-1,\widetilde{j}_2,1,\dots,1} - x_1\beta_1, \alpha^{(2)}_{\widetilde{j}_1,\widetilde{j}_2-1,1,\dots,1} - x_2\beta_2\right).
\end{aligned}$$

If $j_p > 1$ for some $p = 3, \dots, D$, then

$$P\left((x_1\beta_1 + \varepsilon_1, \dots, x_D\beta_D + \varepsilon_D) \in R_{\widetilde{j},\widetilde{j}_2,j_3,\dots,j_D} | x\right) \to 0.$$

Thus,

$$P(Y^{c_1} \le y_{j_1}^{(1)} \,|\, x) \to \sum_{\widetilde{j}_1=1}^{j_1} \sum_{j_2=1}^{M_2} \Big(F_{1,2}\left(\alpha^{(1)}_{\widetilde{j}_1, j_2, 1\ldots,1} - x_1\beta_1, \alpha^{(2)}_{\widetilde{j}_1, j_2, 1\ldots,1} - x_2\beta_2\right) -$$
$$F_{1,2}\left(\alpha^{(1)}_{\widetilde{j}_1-1, j_2, 1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{\widetilde{j}_1, j_2, 1,\ldots,1} - x_2\beta_2\right) -$$
$$F_{1,2}\left(\alpha^{(1)}_{\widetilde{j}_1, j_2, 1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{\widetilde{j}_1, j_2-1, 1,\ldots,1} - x_2\beta_2\right)$$
$$F_{1,2}\left(\alpha^{(1)}_{\widetilde{j}_1-1, j_2, 1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{\widetilde{j}_1, j_2-1, 1,\ldots,1} - x_2\beta_2\right) \Big)$$
$$\text{as} \quad x_{3,1} \to -\infty, \ldots, x_{D,1} \to -\infty.$$

Since all the covariates in x_2 are exclusive, we can vary covariates in x_1 keeping x_2 fixed. Therefore,

$$\lim_{\substack{x_{3,1}\to-\infty,\ldots \\ x_{D,1}\to-\infty}} P(Y^{c_1} \le y_{j_1}^{(1)} \,|\, (\widetilde{x}_1, x_2, \ldots, x_D)) > \lim_{\substack{x_{3,1}\to-\infty,\ldots \\ x_{D,1}\to-\infty}} P(Y^{c_1} \le y_{j_1}^{(1)} \,|\, (\widetilde{\widetilde{x}}_1, x_2, \ldots, x_D))$$
$$\iff \quad \widetilde{x}_1\beta_1 < \widetilde{\widetilde{x}}_1\beta_1. \quad (28)$$

The only difference from what we had above is that instead of taking $x_{2,1} \to -\infty$, we keep the the whole covariate vector x_2 unchanged when analyzing $P(Y^{c_1} \le y_{j_1}^{(1)} \,|\, (x_1, x_2, \ldots, x_D))$. As discussed above, using the continuity of the first covariate in x_1 and the fact that the coefficient $\beta_{1,1}$ is normalized, we can use the same techniques as in Theorem 4 to show that the system of linear inequalities constructed as in (28) identifies β_1.

Coefficients β_d, $d \ge 2$, are identified using an analogous identification strategy. $\square$

B.5 Proof of Theorem 5.

We start by noting that all $\alpha^{(d)}_{1,1,\ldots,1}$ are identified for any $d = 1, \ldots, D$. Indeed, consider observed probabilities

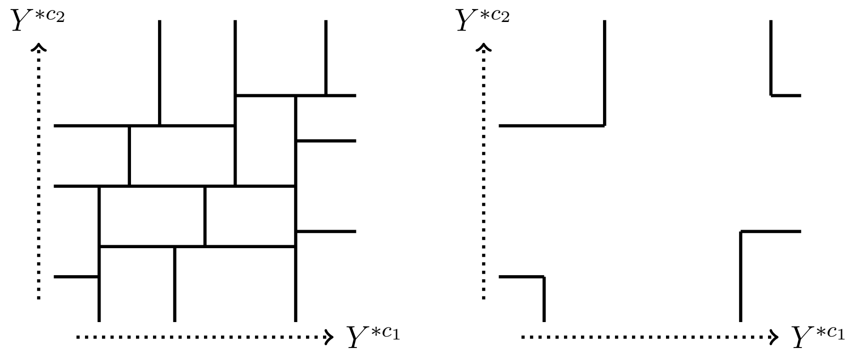
$$P\left(\cap_{h=1}^{D}\left(Y^{c_h} = y_1^{(h)}\right) | \, x\right) = P\left(\cap_h \left(x_h\beta_h + \varepsilon_h \le \alpha^{(h)}_{1,1,\ldots,1}\right) |x\right)$$

and take $x_{h,1} \to -\infty$ for all $h \ne d$. By doing this, we identify $F_d\left(\alpha^{(d)}_{1,1,\ldots,1} - x_d\beta_d\right)$. Now, using the large support condition on $x_{d,1}$, we obtain that $\alpha^{(d)}_{1,1,\ldots,1} - x_d\beta_d$ goes through the whole support of ε_d. Then using the normalization stated in the theorem we find x_{0d} such that

$$F_d\left(\alpha^{(d)}_{1,1,\ldots,1} - x_{0d}\beta_d\right) = c_{0d}$$

and, therefore, we can identify $\alpha^{(d)}_{1,1,\ldots,1}$ as $\alpha^{(d)}_{1,1,\ldots,1} = e_{0d} + x_{0d}\beta_d$ (e_{0d} and β_d are known).

Figure 15: Threshold system we aim to identify (left) and what we identify after Step 1 (right).



Combining the knowledge of $\alpha^{(d)}_{1,1,\ldots,1}$ for any $d = 1, \ldots, D$, with the exclusiveness of some covariates in each index and large support conditions on $x_{d,1}$, $d + 1, \ldots, D$, (it can take any value on the real line), we can identify the joint c.d.f $F(\cdot, \ldots, \cdot)$ from the observed probabilities

$$P\left(\bigcap_{h=1}^{D}\left(Y^{c_h} = y_1^{(h)}\right) \mid x\right) = F\left(\alpha^{(1)}_{1,1,\ldots,1} - x_1\beta_1, \ldots, \alpha^{(D)}_{1,1,\ldots,1} - x_D\beta_D\right).$$

notice that we could have obtained identification of F from any "corner" outcome $(j_1, \ldots, j_D)$, where $j_h \in \{1, M_h\}$ for any $h = 1, \ldots, D$. $\square$

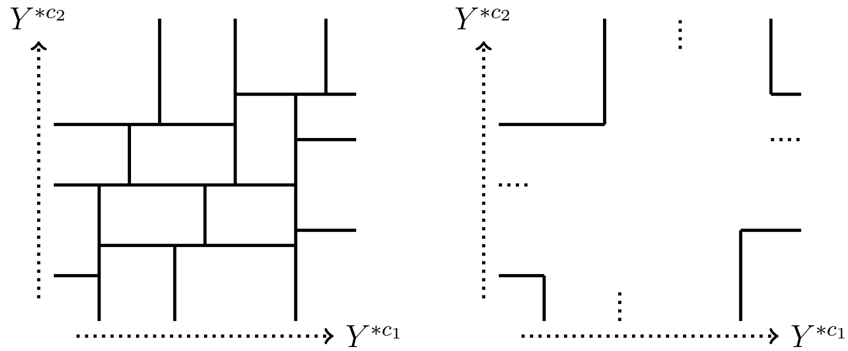
B.6 Proof of Theorem 6.

The proof proceeds to identify all the thresholds sequentially.

Step 1.(**identification of "corner" thresholds**) In the proof of Theorem 5 we already established that $\alpha^{(d)}_{1,1,\ldots,1}$ are identified for any $d = 1, \ldots, D$. In an analogous way we can establish the identification of all the "corner" thresholds $\alpha^{(d)}_{j_1,j_2,\ldots,j_D}$, where $j_d \in \{1, M_d - 1\}$ and for $h \neq d$ each $j_h \in \{1, M_h\}$ (recall that $\alpha^{(d)}_{j_1,\ldots,j_d,\ldots,j_D} = +\infty$ when $j_d = M_d$). The state of what thresholds are identified after this step is illustrated in Figure 15.

Step 2. We now want to show that thresholds $\alpha^{(d)}_{1,1,\ldots,1,j_d,1,\ldots,1}$, $j_d = 1, \ldots, M_d$, where in the own dimension d the discrete response can be any whereas in all the other dimensions the responses are fixed at their lowest values, are identified. These thresholds will be identified from the exclusivity of at

FIGURE 16: Threshold system we aim to identify (left) and what we identify after Step 2 (right).



least one covariate in x_d and large support conditions for that covariate. Indeed, consider

$$P\left(Y^{c_d}=y^{(d)}_{j_d},\cap_{h\neq d}\left(Y^{c_h}=y^{(h)}_1\right)\mid x\right)=P\Bigg(\alpha^{(d)}_{1,1,\dots,1,j_d-1,1,\dots,1}<x_d\beta_d+\varepsilon_d\leq\alpha^{(d)}_{1,1,\dots,1,j_d,1,\dots,1},$$
$$\bigcap_{h\neq d}\left(x_h\beta_h+\varepsilon_h\leq\alpha^{(h)}_{1,1,\dots,1,j_d,1,\dots,1}\right)\Bigg).$$

By taking $x_{h,1}\to-\infty$ for all $h\neq d$, we identify

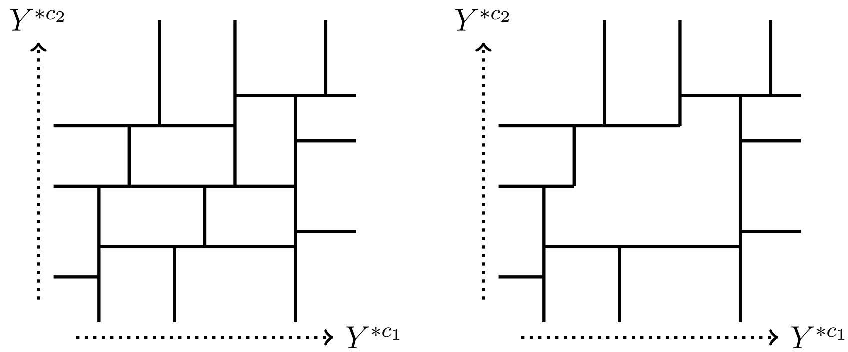
$$\lim_{x_{h,1}\to-\infty,h\neq d}P\left(Y^{c_d}=y^{(d)}_{j_d}\bigcap_{h\neq d}\left(Y^{c_h}=y^{(h)}_1\right)\mid x\right)=$$
$$F_d\big(\alpha^{(d)}_{1,1,\dots,1,j_d,1,\dots,1}-x_d\beta_d\big)-F_d\big(\alpha^{(d)}_{1,1,\dots,1,j_d-1,1,\dots,1}-x_d\beta_d\big)$$

When $j_d=1$, we identify $F_d\big(\alpha^{(d)}_{1,1,\dots,1,j_d,1,\dots,1}-x_d\beta_d\big)$ (this can be seen either from the normalization $\alpha^{(d)}_{1,1,\dots,1,0,1,\dots,1}=-\infty$). When considering $j_d\geq 2$ we can therefore conclude that any $F_d\big(\alpha^{(d)}_{1,1,\dots,1,j_d,1,\dots,1}-x_d\beta_d\big)$ is identified. Since F_d is known from Theorem 5, then by choosing any x_d such that $F_d\big(\alpha^{(d)}_{1,1,\dots,1,j_d,1,\dots,1}-x_d\beta_d\big)\in(0,1)$, we immediately identify $\alpha^{(d)}_{1,1,\dots,1,j_d,1,\dots,1}$.

Here we started with the "corner" of all first responses and allowed responses in one dimension vary. Analogously, we could start with other "corners" and identify all the thresholds $\alpha^{(d)}_{j_1,j_2,\dots,j_{d-1},j_d,j_{d+1},\dots,j_D}$, where $j_d=1,\dots,M_d$ and $j_h\in\{1,M_h\}$, $h\neq d$. The identification after this step is illustrated in Figure 16. The thresholds we identify at this step are in short dot line because we actually don't know the actual respective rectangles yet so we don't know how far these thresholds extend.

Step 3. Now let us show that for any d, any $j_d=1,\dots,M_d$, and any $h_0\neq d$, the threshold $\alpha^{(h_0)}_{1,1,\dots,1,\underbrace{j_d}_{d-th\ position},1,\dots,1}$ is identified. Thus, we consider a threshold in dimension h_0 but allow the response

FIGURE 17: Threshold system we aim to identify (left) and what we identify after Step 3 (right).



in some other dimension to be any. To show this, consider

$$P\left(Y^{c_d}=y^{(d)}_{j_d},\bigcap_{h\neq d}\left(Y^{c_h}=y^{(h)}_1\right)\mid x\right)=P\Bigg(\alpha^{(d)}_{1,1,\dots,1,j_d-1,1,\dots,1}<x_d\beta_d+\varepsilon_d\leq\alpha^{(d)}_{1,1,\dots,1,j_d,1,\dots,1},$$
$$x_{h_0}\beta_{h_0}+\varepsilon_{h_0}\leq\alpha^{(h_0)}_{1,1,\dots,1,j_d,1,\dots,1},\bigcap_{h\neq d,h\neq h_0}\left(x_h\beta_h+\varepsilon_h\leq\alpha^{(h)}_{1,1,\dots,1,j_d,1,\dots,1}\right)\Bigg)$$

and take $x_{h,1}\to-\infty$ for all $h\neq d$, $h\neq h_0$. Then in this limit we identify

$$\begin{aligned}P\left(\alpha^{(d)}_{1,1,\dots,1,j_d-1,1,\dots,1}<x_d\beta_d+\varepsilon_d\leq\alpha^{(d)}_{1,1,\dots,1,j_d,1,\dots,1},x_{h_0}\beta_{h_0}+\varepsilon_{h_0}\leq\alpha^{(h_0)}_{1,1,\dots,1,j_d,1,\dots,1}\right)&=\\ =F_{d,h_0}\left(\alpha^{(d)}_{1,1,\dots,1,j_d,1,\dots,1}-x_d\beta_d,\alpha^{(h_0)}_{1,1,\dots,1,j_d,1,\dots,1}-x_{h_0}\beta_{h_0}\right)&\\ -F_{d,h_0}\left(\alpha^{(d)}_{1,1,\dots,1,j_d-1,1,\dots,1}-x_d\beta_d,\alpha^{(h_0)}_{1,1,\dots,1,j_d,1,\dots,1}-x_{h_0}\beta_{h_0}\right),&\end{aligned}\tag{29}$$

where F_{d,h_0} is the joint c.d.f. of $(\varepsilon_d,\varepsilon_{h_0})$. This c.d.f. is already identified from Theorem 5, and $\alpha^{(d)}_{1,1,\dots,1,j_d,1,\dots,1}-x_d\beta_d$ and $\alpha^{(d)}_{1,1,\dots,1,j_d-1,1,\dots,1}-x_d\beta_d$ are already identified too.

Note that for known e_1, $\Delta e_1>0$, the function

$$F_{d,h_0}\left(e_1+\Delta e_1,e_2\right)-F_{d,h_0}\left(e_1,e_2\right)$$

is known as a function of e_2 and is strictly increasing in e_2 (if, of course, both $e_1+\Delta e_1$ and e_1 are in the support of ε_d). Therefore, from the known probability on the left-hand side of 29, we can immediately identify $\alpha^{(h_0)}_{1,1,\dots,1,j_d,1,\dots,1}$.

Of course, an analogous proof would apply when instead of some of the 1's in $\alpha^{(h_0)}_{1,1,\dots,1,j_d,1,\dots,1}$ we have the highest values M_d's, thus effectively considering different "corners". The identification after this step is illustrated in Figure 17.

Step 4. Now let us show that for any d, any $j_d = 1, \ldots, M_d$, and any $h_0 \neq d$, the threshold $\alpha^{(h_0)}_{1,1,\ldots,1,\underbrace{j_{h_0}}_{h_0-th\ position},1,\ldots,1,\underbrace{j_d}_{d-th\ position},1,\ldots,1}$ is identified. In step 3 we already established it for $j_{h_0} = 1$. Thus, for a given dimension, we allow the discrete responses in that dimension and some other dimension to be arbitrary.

To make notations a bit simpler, we will suppose that $d = 2$ and $h_0 = 1$ and, thus, prove that any threshold $\alpha^{(1)}_{j_1,j_2,1,\ldots,1}$ as well as $\alpha^{(2)}_{j_1,j_2,1,\ldots,1}$ is identified. We can identify these thresholds sequentially starting from one of the "corners" in the first two dimensions.

If, for example, we start from the bottom left "corner" we will first take $j_1 = 2$ and $j_2 = 2$ and consider the following observed probability:

$$P\left(Y^{c_1} = y^{(1)}_2, Y^{c_d} = y^{(2)}_2, \bigcap_{h\neq 2, h\neq 1}\left(Y^{c_h} = y^{(h)}_1\right) | \, x\right) = P\Big(\alpha^{(1)}_{1,1,\ldots,1,1,1,\ldots,1} < x_1\beta_1 + \varepsilon_1 \leq \alpha^{(1)}_{2,2,1,\ldots,1},$$
$$\alpha^{(2)}_{2,1,1,\ldots,1} < x_2\beta_2 + \varepsilon_2 \leq \alpha^{(2)}_{2,2,1,\ldots,1}, \bigcap_{h\neq d, h\neq 1}\left(x_h\beta_h + \varepsilon_h \leq \alpha^{(h)}_{2,2,1,\ldots,1}\right) | \, x\Big).$$

Taking $x_{h,1} \to -\infty$ for all $h \neq 2$, $h \neq 1$, in the limit we known

$$\begin{aligned}
&P\left(\alpha^{(1)}_{1,2,1,\ldots,1} < x_1\beta_1 + \varepsilon_1 \leq \alpha^{(1)}_{2,2,1,\ldots,1}, \alpha^{(2)}_{2,1,1,\ldots,1} < x_2\beta_2 + \varepsilon_2 \leq \alpha^{(2)}_{2,2,1,\ldots,1} | \, x\right) = \\
&\quad = F_{1,2}\left(\alpha^{(1)}_{2,2,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{2,2,1,\ldots,1} - x_2\beta_2\right) - F_{1,2}\left(\alpha^{(1)}_{2,2,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{2,1,1,\ldots,1} - x_2\beta_2\right) - \\
&\qquad F_{1,2}\left(\alpha^{(1)}_{1,2,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{2,2,1,\ldots,1} - x_2\beta_2\right) + F_{1,2}\left(\alpha^{(1)}_{1,2,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{2,1,1,\ldots,1} - x_2\beta_2\right).
\end{aligned} \tag{30}$$

On the right-hand side in (30) we have a known function $F_{1,2}$ and known $\alpha^{(1)}_{1,2,1,\ldots,1}$ and $\alpha^{(2)}_{2,1,1,\ldots,1}$ (they were identified in previous steps). In some situations one of the parameters $\alpha^{(1)}_{2,2,1,\ldots,1}$ or $\alpha^{(2)}_{2,2,1,\ldots,1}$ may be known. E.g., in the situation described in the right-hand side of Figure 17 when trying to identify the rectangle corresponding to the response $(y^{(1)}_2, y^{(2)}_2)$, the threshold $\alpha^{(2)}_{2,2,1,\ldots,1}$ is known as well, thus, giving three known sides of the respective rectangle. In this case, the identification can proceed in the same way way as in Step 3 as the right-hand side in (30) has only one unknown parameter $\alpha^{(1)}_{2,2,1,\ldots,1}$ and is strictly monotone in that parameter when we choose x such that the probability on the left-hand side of (30) is strictly between 0 and 1. However, we also need a strategy for the case when both $\alpha^{(1)}_{2,2,1,\ldots,1}$ or $\alpha^{(2)}_{2,2,1,\ldots,1}$ may be unknown at this stage. Indeed, this would be analogous to the situation described in the right-hand side of Figure 17 when trying to identify the rectangle corresponding to the response $(y^{(1)}_3, y^{(2)}_2)$ – in that case only two sides of the rectangle are already identified. The problem of identifying both $\alpha^{(1)}_{2,2,1,\ldots,1}$ or $\alpha^{(2)}_{2,2,1,\ldots,1}$ can be reformulated as showing that there is only one set of parameters $(\Delta_{1A}, \Delta_{2A})$, $\Delta_{iA} > 0$, $i = 1, 2$, such that for all (z_1, z_2)

$$Q_{1,2}(z_1, z_2) = F_{1,2}(\Delta_{1A} + z_1, \Delta_{2A} + z_2) - F_{1,2}(\Delta_{1A} + z_1, z_2) - F_{1,2}(z_1, \Delta_{2A} + z_2) + F_{1,2}(z_1, z_2), \tag{31}$$

where $Q_{1,2}(z_1, z_2)$ is known and denotes, of course, the probability of choice. Clearly, $z_1 = \alpha^{(1)}_{1,2,1,\ldots,1} - x_1\beta_1$ and $z_2 = \alpha^{(2)}_{2,1,1,\ldots,1} - x_2\beta_2$, whereas by Δ_{1A} and Δ_{2A} we mean $\Delta_{1A} = \alpha^{(1)}_{2,2,1,\ldots,1} - \alpha^{(1)}_{1,2,1,\ldots,1} > 0$, $\Delta_{2A} = \alpha^{(2)}_{2,2,1,\ldots,1} - \alpha^{(2)}_{2,1,1,\ldots,1} > 0$.

If, for example, we start from the upper left "corner" we will first take $j_1 = 2$ and $j_2 = M_2 - 2$ and consider the following observed probability:

$$P\left(Y^{c_1} = y^{(1)}_2, Y^{c_d} = y^{(2)}_{M_2-1}, \bigcap_{h\neq 2, h\neq 1}\left(Y^{c_h} = y^{(h)}_1\right) | \, x\right) = P\Big(\alpha^{(1)}_{1,M_2-1,\ldots,1,1,1,\ldots,1} < x_1\beta_1 + \varepsilon_1 \le \alpha^{(1)}_{2,M_2-1,1,\ldots,1},$$
$$\alpha^{(2)}_{2,M_2-2,1,\ldots,1} < x_2\beta_2 + \varepsilon_2 \le \alpha^{(2)}_{2,M_2-1,1,\ldots,1}, \bigcap_{h\neq d, h\neq 1}\left(x_h\beta_h + \varepsilon_h \le \alpha^{(h)}_{2,M_2-1,1,\ldots,1}\right) | \, x\Big).$$

Taking $x_{h,1} \to -\infty$ for all $h \neq 2$, $h \neq 1$, in the limit we known

$$\begin{aligned} &P\left(\alpha^{(1)}_{1,M_2-1,1,\ldots,1} < x_1\beta_1 + \varepsilon_1 \le \alpha^{(1)}_{2,M_2-1,1,\ldots,1}, \alpha^{(2)}_{2,M_2-2,1,\ldots,1} < x_2\beta_2 + \varepsilon_2 \le \alpha^{(2)}_{2,M_2-1,1,\ldots,1} | \, x\right) = \\ &= F_{1,\overline{2}}\left(\alpha^{(1)}_{2,M_2-1,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{2,M_2-2,1,\ldots,1} - x_2\beta_2\right) - F_{1,\overline{2}}\left(\alpha^{(1)}_{2,M_2-1,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{2,M_2-1,1,\ldots,1} - x_2\beta_2\right) - \\ &F_{1,\overline{2}}\left(\alpha^{(1)}_{1,M_2-1,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{2,M_2-2,1,\ldots,1} - x_2\beta_2\right) + F_{1,\overline{2}}\left(\alpha^{(1)}_{1,M_2-1,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{2,M_2-1,1,\ldots,1} - x_2\beta_2\right). \end{aligned} \tag{32}$$

On the right-hand side in (32) we have a known function $F_{1,\overline{2}}$ and known $\alpha^{(1)}_{1,M_2-1,1,\ldots,1}$ and $\alpha^{(2)}_{2,M_2-1,1,\ldots,1}$ (they were identified in previous steps). In some situations one of the parameters $\alpha^{(1)}_{2,M_2-1,1,\ldots,1}$ or $\alpha^{(2)}_{2,M_2-2,1,\ldots,1}$ may be known and then the other parameter would be easy t identify from the monotonicity properties of $F_{1,\overline{2}}$. However, we need a strategy for cases when both these parameters may be unknown at this stage. The problem of identifying both these parameters can be reformulated as showing that there is only one set of parameters $(\Delta_{1B}, \Delta_{2B})$, $\Delta_{1B} > 0$, $\Delta_{2B} < 0$, such that for all (z_1, z_2)

$$Q_{1,\overline{2}}(z_1, z_2) = F_{1,\overline{2}}(\Delta_{1B} + z_1, \Delta_{2B} + z_2) - F_{1,\overline{2}}(\Delta_{1B} + z_1, z_2) - F_{1,\overline{2}}(z_1, \Delta_{2B} + z_2) + F_{1,\overline{2}}(z_1, z_2), \tag{33}$$

where $Q_{1,\overline{2}}(z_1, z_2)$ is known and denotes, of course, the probability of choice. Clearly, $z_1 = \alpha^{(1)}_{1,M_2-1,1,\ldots,1} - x_1\beta_1$ and $z_2 = \alpha^{(2)}_{2,M_2-1,1,\ldots,1} - x_2\beta_2$, whereas by Δ_{1B} and Δ_{2B} we mean $\Delta_{1B} = \alpha^{(1)}_{2,M_2-1,1,\ldots,1} - \alpha^{(1)}_{1,M_2-1,1,\ldots,1} > 0$ and $\Delta_{2B} = \alpha^{(2)}_{2,M_2-2,1,\ldots,1} - \alpha^{(2)}_{2,M_2-1,1,\ldots,1} < 0$.

If, for example, we start from the bottom right "corner" we will first take $j_1 = M_1 - 2$ and $j_2 = 2$ and consider the following observed probability:

$$P\left(Y^{c_1} = y^{(1)}_{M_1-1}, Y^{c_d} = y^{(2)}_2, \bigcap_{h\neq 2, h\neq 1}\left(Y^{c_h} = y^{(h)}_1\right) | \, x\right) = P\Big(\alpha^{(1)}_{M_1-2,2,\ldots,1,1,1,\ldots,1} < x_1\beta_1 + \varepsilon_1 \leq \alpha^{(1)}_{M_1-1,2,1,\ldots,1},$$

$$\alpha^{(2)}_{M_1-1,1,1,\ldots,1} < x_2\beta_2 + \varepsilon_2 \leq \alpha^{(2)}_{M_1-1,2,1,\ldots,1}, \bigcap_{h\neq d, h\neq 1}\left(x_h\beta_h + \varepsilon_h \leq \alpha^{(h)}_{M_1-1,2,1,\ldots,1}\right) | \, x\Big).$$

Taking $x_{h,1} \to -\infty$ for all $h \neq 2$, $h \neq 1$, in the limit we known

$$P\Big(\alpha^{(1)}_{M_1-2,2,\ldots,1,1,1,\ldots,1} < x_1\beta_1 + \varepsilon_1 \leq \alpha^{(1)}_{M_1-1,2,1,\ldots,1}, \alpha^{(2)}_{M_1-1,1,1,\ldots,1} < x_2\beta_2 + \varepsilon_2 \leq \alpha^{(2)}_{M_1-1,2,1,\ldots,1} | \, x\Big) =$$
$$= F_{\bar{1},2}\left(\alpha^{(1)}_{M_1-2,2,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{M_1-1,2,1,\ldots,1} - x_2\beta_2\right) - F_{\bar{1},2}\left(\alpha^{(1)}_{M_1-1,2,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{M_1-1,2,1,\ldots,1} - x_2\beta_2\right) -$$
$$F_{\bar{1},2}\left(\alpha^{(1)}_{M_1-2,2,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{M_1-1,1,1,\ldots,1} - x_2\beta_2\right) + F_{\bar{1},2}\left(\alpha^{(1)}_{M_1-1,2,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{M_1-1,1,1,\ldots,1} - x_2\beta_2\right). \tag{34}$$

On the right-hand side in (34) we have a known function $F_{\bar{1},2}$ and known $\alpha^{(1)}_{M_1-1,2,1,\ldots,1}$ and $\alpha^{(2)}_{M_1-1,1,1,\ldots,1}$ (they were identified in previous steps). In some situations one of the parameters $\alpha^{(1)}_{M_1-2,2,1,\ldots,1}$ or $\alpha^{(2)}_{M_1-1,2,1,\ldots,1}$ may be known and then the other parameter would be easy to identify from the monotonicity properties of $F_{\bar{1},2}$. However, we need a strategy for cases when both these parameters may be unknown at this stage. The problem of identifying both these parameters can be reformulated as showing that there is only one set of parameters $(\Delta_{1C}, \Delta_{2C})$, $\Delta_{1C} < 0$, $\Delta_{2C} > 0$, such that for all (z_1, z_2)

$$Q_{\bar{1},2}(z_1, z_2) = F_{\bar{1},2}(\Delta_{1C} + z_1, \Delta_{2C} + z_2) - F_{\bar{1},2}(\Delta_{1C} + z_1, z_2) - F_{\bar{1},2}(z_1, \Delta_{2C} + z_2) + F_{\bar{1},2}(z_1, z_2), \tag{35}$$

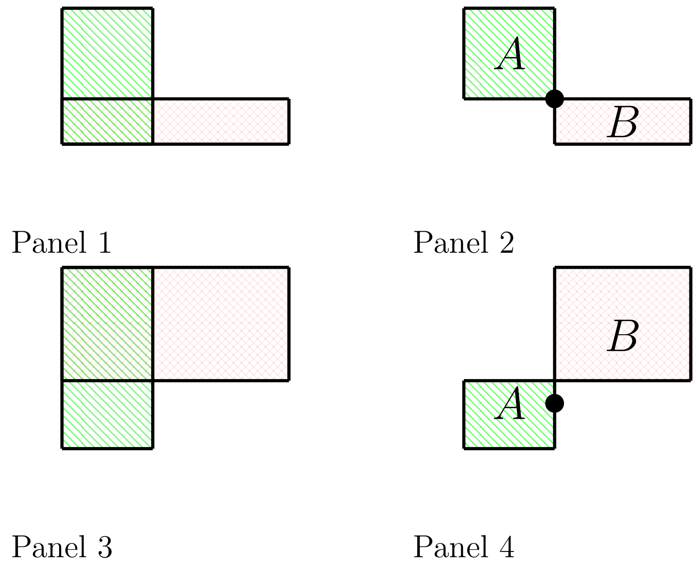
where $Q_{\bar{1},2}(z_1, z_2)$ is known and denotes, of course, the probability of choice. Clearly, $z_1 = \alpha^{(1)}_{M_1-1,2,1,\ldots,1} - x_1\beta_1$ and $z_2 = \alpha^{(2)}_{M_1-1,1,1,\ldots,1} - x_2\beta_2$, whereas by Δ_{1C} and Δ_{2C} we mean $\Delta_{1C} = \alpha^{(1)}_{M_1-2,2,1,\ldots,1} - \alpha^{(1)}_{M_1-1,2,1,\ldots,1} < 0$ and $\Delta_{2C} = \alpha^{(2)}_{M_1-1,2,1,\ldots,1} - \alpha^{(2)}_{M_1-1,1,1,\ldots,1} > 0$.

Analogously, we can consider the remaining upper right "corner".

For now, suppose we started identification with the bottom left "corner". Suppose that there is another set of parameters $(\delta_{1A}, \delta_{2A})$, $\delta_{iA} > 0$, $i = 1, 2$, such that for all (z_1, z_2)

$$Q_{1,2}(z_1, z_2) = F_{1,2}(\delta_{1A} + z_1, \delta_{2A} + z_2) - F_{1,2}(\delta_{1A} + z_1, z_2) - F_{1,2}(z_1, \delta_{2A} + z_2) + F_{1,2}(z_1, z_2).$$

Figure 18: First illustration of identification in Step 4.



Notes: Panel 1: two overlapping rectangles have the same probability mass when identification proceeds from the bottom left "corner". Panel 2: two non-overlapping rectangles have the same probability mass when identification proceeds from the bottom left "corner". Panel 3: two overlapping rectangles have the same probability mass when identification proceeds from the upper left "corner". Panel 4: two non-overlapping rectangles have the same probability mass when identification proceeds from the upper left "corner".

This necessarily implies that $(\Delta_{1A} - \delta_{1A})(\Delta_{2A} - \delta_{2A}) < 0$. The condition that both these sets of parameters give the same observed probabilities $Q_{1,2}(z_1, z_2)$ corresponds to the picture in Panel 1 in Figure 18, where the red rectangle and the green rectangle have to contain equal probability mass. By removing the shared rectangle, we conclude that the two rectangles in Panel 2 in Figure 18, which are modified versions of the two rectangles on the left-hand side, have to contain equal probability mass. We will call the point that joins two rectangles on the right-hand side diagram in Figure 18 as the *join*. It is depicted on the right panel of Figure 18 as a thick dot.

If we suppose that we started identification with the upper left "corner". Suppose that there is another set of parameters $(\delta_{1B}, \delta_{2B})$, $\delta_{1B} > 0$, $\delta_{2B} < 0$, such that for all (z_1, z_2)

$$Q_{1,\overline{2}}(z_1, z_2) = F_{1,2}(\delta_{1B} + z_1, \delta_{2B} + z_2) - F_{1,2}(\delta_{1B} + z_1, z_2) - F_{1,2}(z_1, \delta_{2B} + z_2) + F_{1,2}(z_1, z_2).$$

Of course, this necessarily implies that $(\Delta_{1B} - \delta_{1B})(|\Delta_{2B}| - |\delta_{2B}|) < 0$. The condition that both these sets of parameters give the same observed probabilities $Q_{1,\overline{2}}(z_1, z_2)$ corresponds to the picture in Panel 3 in Figure 18, where the red rectangle and the green rectangle have to contain equal probability mass. By removing the shared rectangle, we conclude that the two rectangles in Panel 4 in Figure 18, which are modified versions of the two rectangles on the left-hand side, have to contain equal probability mass.

Let is denote the support of $(\varepsilon_1, \varepsilon_2)$ as $\mathcal{E}_{12}$.

(i) Consider first the case when $\mathcal{E}_{12}$ has an extreme point which is also a global optimum in some dimension. Suppose, e.g., that there is an extreme point whose first coordinate is aa global maximum in the first dimension. Take such an extreme point as the join point. Then by construction, the probability mass of rectangular region analogous to B in Figure 18 is zero. If the rectangular region analogous to A in Figure 18 has an overlap with $\mathcal{E}_{12}$ and this overlap has a non-empty interior, we obtain a contradiction as the probability mass of region A is strictly positive, whereas the probability mass of B is zero. This situation is illustrated in Panel 1 of Figure 19, where the gray area is a part of $\mathcal{E}_{12}$ around the node point of interest (we want to note here that the boundary depicted as straight line in the graph is only drawn in this way for illustrational simplicity; in general, of course, the boundary curve will not necessarily be straight but will be that of a general convex set – however, the same argument will apply to such a general case). One can see that this corresponds to the situation when the join point corresponds to the global maximum value of $\mathcal{E}_{12}$ in the first dimension but in the second dimension it is not a global maximum.

It is possible that the probability mass of A is zero but this means that we are in the situation when the join point corresponds to the global maximum value of $\mathcal{E}_{12}$ in both the first and the second dimensions. These situations are illustrated in Panel 2, Panel 4 and Panel 6 of Figure 19. In all these situation we can move the join along the boundary in the clockwise direction (see Panel 3, Panel 5 and Panel 7 in Figure 19) and obtain a situation with different A and B but now A has a strictly positive probability mass whereas B has a zero mass.

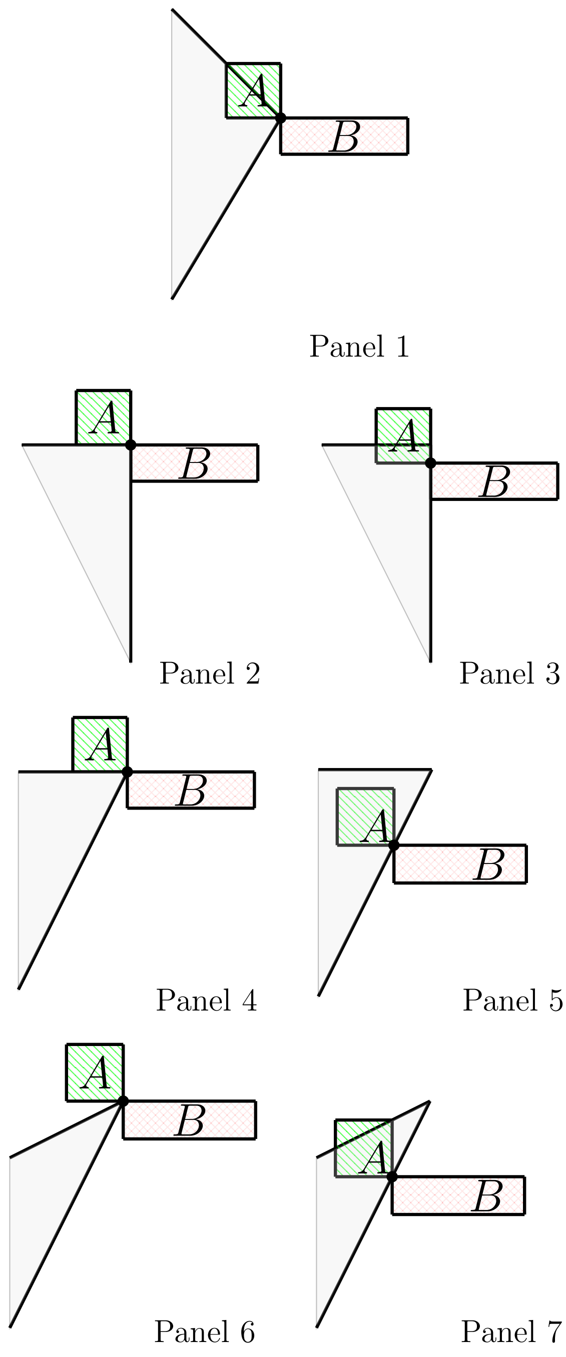
Our discussion so far obtains a contradiction for one point $(z_1, z_2) = (\alpha^{(1)}_{1,2,1,\ldots,1} - x_1\beta_1, \alpha^{(2)}_{2,1,1,\ldots,1} - x_2\beta_2)$ on the boundary of the support $\mathcal{E}_{12}$. It implies though that the contradiction can also be obtained also for a strictly positive mass of (z_1, z_2) in $\mathcal{E}_{12}$ (hence, a strictly positive mass of (x_1, x_2)) in the neighborhood of this boundary point. Indeed, the important implication of the above constructions (under the supposition of (Δ_1, Δ_2) and (δ_1, δ_2) being both observationally equivalent) is the discontinuity in the probability masses of the described regions A and B. This discontinuity and, hence, the contradiction will remain if instead of (z_1, z_2) we consider points in $\mathcal{E}_{12}$ that are in a neighborhood of (z_1, z_2).

Cases when an extreme point attains a global minimum in the first dimension or attains a global optimum in the second dimension are considered analogously.

(ii) Let us now consider case when the interior of $\mathcal{E}_{12}$ contains points that in each coordinate are unbounded either from above or from below.

First, consider the case when for some (z_{10}, z_{20}) the interior of $\mathcal{E}_{12}$ contains all the points in the quadrant

Figure 19: Second illustration of identification in Step 4



$(z_{10}, z_{20}) + \mathcal{O}_{++}$, where

$$\mathcal{O}_{s_1 s_2} = \{(s_1\lambda_1, s_2\lambda_2) : \lambda_i \geq 0, i = 1, 2\}.$$

In this case we will use the identification strategy for thresholds that starts with the bottom left "corner". As mentioned above, we should have either $\Delta_{1A} > \delta_{1A}$ or $\Delta_{2A} > \delta_{2A}$. Suppose that $\Delta_{1A} > \delta_{1A}$ (then, necessarily, $\Delta_{2A} < \delta_{2A}$).

Note that for the fixed point $(z_{10,\lambda_1}, z_{20,\lambda_2})$, where $z_{i0,\lambda_i} = z_{i0} + \lambda_i$, $i = 1, 2$, the relation

$$Q_{1,2}\left(z_{10,\lambda_1}, z_{20,\lambda_2}\right) = F_{1,2}\left(\tilde{\Delta}_1 + z_{10,\lambda_1}, \tilde{\Delta}_2 + z_{20,\lambda_2}\right) - F_{1,2}\left(\tilde{\Delta}_1 + z_{10,\lambda_1}, z_{20,\lambda_2}\right) \\ - F_{12}\left(z_{10,\lambda_1}, \tilde{\Delta}_2 + z_{20,\lambda_2}\right) + F_{12}\left(z_{10,\lambda_1}, z_{20,\lambda_2}\right) \quad (36)$$

describes a decreasing function $\psi_{z_{10,\lambda_1}, z_{20,\lambda}}(\cdot)$ such that

$$Q_{1,2}\left(z_{10,\lambda_1}, z_{20,\lambda_2}\right) = F_{1,2}\left(\tilde{\Delta}_1 + z_{10,\lambda_1}, \psi_{z_{10,\lambda_1}, z_{20,\lambda_2}}(\tilde{\Delta}_1) + z_{20,\lambda_2}\right) - F_{1,2}\left(\tilde{\Delta}_1 + z_{10,\lambda_1}, z_{20,\lambda_2}\right) \\ - F_{1,2}\left(z_{10,\lambda_1}, \psi_{z_{10,\lambda_1}, z_{20,\lambda_2}}(\tilde{\Delta}_1) + z_{20,\lambda_2}\right) + F_{1,2}\left(z_{10,\lambda_1}, z_{20,\lambda_2}\right).$$

Of course, we have that $\Delta_{2A} = \psi_{z_{10,\lambda_1}, z_{20,\lambda_2}}(\Delta_{1A})$ and by our supposition that $(\delta_{1A}, \delta_{2A})$ can rationalize the data as well, we have that $\delta_{2A} = \psi_{z_{10,\lambda_1}, z_{20,\lambda_2}}(\delta_{1A})$. We note that $\psi_{z_{10,\lambda_1}, z_{20,\lambda_2}}$ is strictly decreasing is obvious as the right-hand side of (36) is strictly increasing in $\tilde{\Delta}_1$ and is strictly increasing in $\tilde{\Delta}_2$. Note, however, that $\psi_{z_{10,\lambda_1}, z_{20,\lambda_2}}(\cdot)$ is defined on $(\underline{\tilde{\Delta}_1}(z_{10,\lambda_1}, z_{20,\lambda_2}), +\infty)$, where the infimum point corresponds to the case when $\psi_{z_{10,\lambda_1}, z_{20,\lambda_2}}(\underline{\tilde{\Delta}_1}) = +\infty$, and, thus, can be defined as the solution to the following correspondence:

$$Q_{1,2}\left(z_{10,\lambda_1}, z_{20,\lambda_2}\right) = F_1\left(\underline{\tilde{\Delta}_1}(z_{10,\lambda_1}, z_{20,\lambda_2}) + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right) \\ - F_{1,2}\left(\underline{\tilde{\Delta}_1}(z_{10,\lambda_1}, z_{20,\lambda_2}) + z_{10,\lambda_1}, z_{20,\lambda_2}\right) + F_{1,2}\left(z_{10,\lambda_1}, z_{20,\lambda_2}\right),$$

which, of course, describes an unbounded orange region illustrated in Figure 20. By construction, the probability mass of the orange and the blue regions in Figure 20 coincide. We note however, that through the choice of (λ_1, λ_2) we can always make $\underline{\tilde{\Delta}_1}(z_{10,\lambda_1}, z_{20,\lambda_2})$ to get arbitrarily closely to Δ_{1A}. Indeed, for any $\delta_{1A} < \Delta_{1A}$, we can find $\lambda_1, \lambda_2 \geq 0$ such that

$$\frac{F_{12}\left(\Delta_{1A} + z_{10,\lambda_1}, \Delta_{2A} + z_{20,\lambda_2}\right) - F_{12}\left(z_{10,\lambda_1}, \Delta_{2A} + z_{20,\lambda_2}\right) - F_{12}\left(\Delta_{1A} + z_{10,\lambda_1}, z_{20,\lambda_2}\right) + F_{12}\left(z_{10,\lambda_1}, z_{20,\lambda_2}\right)}{F_1\left(\delta_{1A} + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right) - F_{12}\left(\delta_{1A} + z_{10,\lambda_1}, z_{20,\lambda_2}\right) + F_{12}\left(z_{10,\lambda_1}, z_{20,\lambda_2}\right)} > 1, \quad (37)$$

which immediately implies that we should have $\underline{\tilde{\Delta}_1}(z_{10,\lambda_1}, z_{20,\lambda_2}) > \delta_{1A}$ (and as λ_2 can be chosen very large, we can make $\underline{\tilde{\Delta}_1}(z_{10,\lambda_1}, z_{20,\lambda_2})$ arbitrarily close to Δ_{1A}). This gives us a contradiction that parameter δ_{1A} together with δ_{2A} is observationally equivalent to Δ_{1A} and Δ_{2A}, where $\Delta_{1A} > \delta_{1A}$.

Let us now discuss in more detail the claim of being able to choose $\lambda_1, \lambda_2 \geq 0$ such that such that (37) holds. This fact follows from the properties of the bivariate c.d.f. and can be especially easily seen when ε_1 and ε_2 are independent as then (37) can be rewritten as

$$\frac{\left(F_1\left(\Delta_{1A} + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right)\right) \cdot \left(F_2\left(\Delta_{2A} + z_{20,\lambda_2}\right) - F_2\left(z_{20,\lambda_2}\right)\right)}{\left(F_1\left(\delta_{1A} + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right)\right) \cdot \left(1 - F_2\left(z_{20,\lambda_2}\right)\right)},$$

as it is obvious that we can fix λ_1 and take $\lambda_2 \to \infty$, in which case we have $\frac{F_1\left(\Delta_{1A}+z_{10,\lambda_1}\right)-F_1\left(z_{10,\lambda_1}\right)}{F_1\left(\delta_{1A}+z_{10,\lambda_1}\right)-F_1\left(z_{10,\lambda_1}\right)} > 1$ and $\frac{F_2\left(\Delta_{2A}+z_{20,\lambda_2}\right)-F_2\left(z_{20,\lambda_2}\right)}{1-F_2\left(z_{20,\lambda_2}\right)} \to 1$. For a general bivariate c.d.f., in order to show (37) one has to note that a bivariate copula $C(\cdot,\cdot)$ satisfies $\max\{0, u+v-1\}C(u,v) \leq \min\{u,v\}$, which implies that for a fixed u, $|C(uv) - uv| \to 0$ as $v \to 1$, and, hence, for $u_2 > u_1$, one has that

$$C(u_2, v) - C(u_1, v) = (u_2 - u_1)v + (u_2 - u_1) \cdot o(1 - v) \quad \text{as} \quad v \to 1.$$

This observation allows us to rewrite (37) as A_1/A_2, where

$$A_1 = \left(F_1\left(\Delta_1 + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right)\right) \cdot \left(F_2\left(\Delta_2 + z_{20,\lambda_2}\right) - F_2\left(z_{20,\lambda_2}\right)\right)$$
$$+ \left(F_1\left(\Delta_1 + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right)\right) \cdot o(1 - F_2\left(\Delta_2 + z_{20,\lambda_2}\right)) + \left(F_1\left(\Delta_1 + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right)\right) \cdot o(1 - F_2\left(z_{20,\lambda_2}\right)),$$

$$A_2 = \left(F_1\left(\delta_1 + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right)\right) \cdot \left(1 - F_2\left(z_{20,\lambda_2}\right)\right) + \left(F_1\left(\delta_1 + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right)\right) \cdot o(1 - F_2\left(z_{20,\lambda_2}\right)),$$

as $\lambda_2 \to +\infty$. The terms that have the slowest rate of converging to zero as $\lambda_2 \to +\infty$ in the numerator A_1 and the denominator A_2 are the terms

$$\left(F_1\left(\Delta_1 + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right)\right) \cdot \left(F_2\left(\Delta_2 + z_{20,\lambda_2}\right) - F_2\left(z_{20,\lambda_2}\right)\right)$$

and

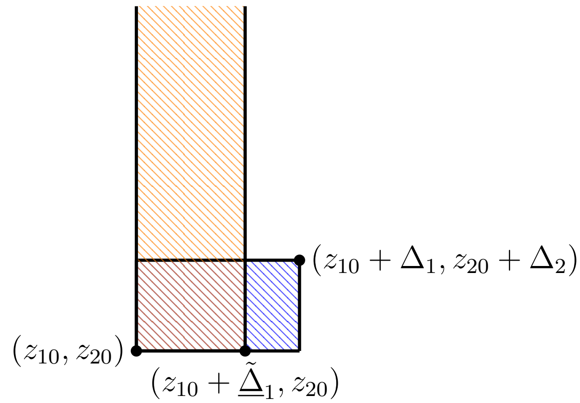
$$\left(F_1\left(\delta_1 + z_{10,\lambda_1}\right) - F_1\left(z_{10,\lambda_1}\right)\right) \cdot \left(1 - F_2\left(z_{20,\lambda_2}\right)\right),$$

respectively. Therefore, the limit of A_1/A_2 as $\lambda_2 \to +\infty$ coincides with the limit of (37) as $\lambda_2 \to +\infty$. This limit is $\frac{F_1\left(\Delta_1+z_{10,\lambda_1}\right)-F_1\left(z_{10,\lambda_1}\right)}{F_1\left(\delta_1+z_{10,\lambda_1}\right)-F_1\left(z_{10,\lambda_1}\right)} > 1$.

To summarize this subcase, by properties of $\mathcal{E}_{12}$ (convexity and non-empty interior) there will be a positive measure of $(z_{10}, z_{20}) \in \mathcal{E}_{12}$ such that $(z_{10}, z_{20}) + \mathcal{O}_{++}$ is contained in $\mathcal{E}_{12}$. Therefore, for a positive measure of $(z_{10,\lambda_1}, z_{20,\lambda_2})$ we obtain that $\underline{\tilde{\Delta}_1}(z_{10,\lambda_1}, z_{20,\lambda_2}) > \delta_{1A}$ giving us the contradiction that parameter δ_{1A} together with δ_{2A} is observationally equivalent to Δ_{1A} and Δ_{2A}, where $\Delta_{1A} > \delta_{1A}$. This contradiction allows us to conclude that $(\Delta_{1A}, \Delta_{2A})$ is identified from (31). Note that we can modify the proof of this subcase by instead considering a fixed λ_2 and taking $\lambda_1 \to +\infty$.

If for some (z_{10}, z_{20}) the interior of $\mathcal{E}_{12}$ contains all the points in the quadrant $(z_{10}, z_{20}) + \mathcal{O}_{+-}$, then we

FIGURE 20: Third illustration of identification in Step 4.



would use the identification strategy for thresholds that starts with the upper left "corner". We would use the function $Q_{1,\overline{2}}(\cdot,\cdot)$ and the points $(z_{10,\lambda_1}, z_{20,\lambda_2})$, where $z_{10,\lambda_1} = z_{i0} + \lambda_1$ and $z_{20,\lambda_2} = z_{20} - \lambda_2$, to obtain a contradiction that there are two different sets of parameters $(\Delta_{1B}, \Delta_{2B})$ and $(\delta_{1B}, \delta_{2B})$ that would give the same function $Q_{1,\overline{2}}(\cdot,\cdot)$.

If for some (z_{10}, z_{20}) the interior of $\mathcal{E}_{12}$ contains all the points in the quadrant $(z_{10}, z_{20}) + \mathcal{O}_{-+}$, then we would use the identification strategy for thresholds that starts with the bottom right "corner". We would use the function $Q_{\overline{1},2}(\cdot,\cdot)$ and the points $(z_{10,\lambda_1}, z_{20,\lambda_2})$, where $z_{10,\lambda_1} = z_{i0} - \lambda_1$ and $z_{20,\lambda_2} = z_{20} + \lambda_2$, to obtain a contradiction that there are two different sets of parameters $(\Delta_{1C}, \Delta_{2C})$ and $(\delta_{1C}, \delta_{2C})$ that would give the same function $Q_{\overline{1},2}(\cdot,\cdot)$.

If for some (z_{10}, z_{20}) the interior of $\mathcal{E}_{12}$ contains all the points in the quadrant $(z_{10}, z_{20}) + \mathcal{O}_{--}$, then in analogous way we would use the identification strategy for thresholds that starts with the upper right "corner".

(iii) Finally, we consider the intermediate case when (a) $\mathcal{E}_{12}$ does not have an extreme point whose coordinate in some dimension is a global extremum of $\mathcal{E}_{12}$ in that dimension, and at the same time (b) $\mathcal{E}_{12}$ does not contain any quadrants in the form $(z_{10}, z_{20}) + \mathcal{O}_{s_1\, s_2}$.

We can establish that in this case $\mathcal{E}_{12}$ is a region between two parallel lines. In other words, $\mathcal{E}_{12}$ can be represented as

$$\mathcal{E}_{12} = \{\underbrace{(z_{10}, z_{20})}_{z_0} + \lambda \cdot (g_1, g_2)\,|\,\lambda \in \mathbb{R}, z_0 \in \mathcal{E}_{12}\}, \quad g_1, g_2 \neq 0, \tag{38}$$

and $\exists z_0^*, z_0^{**} \in \mathcal{E}_{12}$ such that

$$\forall\, \lambda \in \mathbb{R} \qquad \{z_0^* + \lambda \cdot (g_1, g_2) + \mu(g_2, -g_1) : \mu > 0\} \cap \mathcal{E}_{12} = \emptyset, \tag{39}$$

and also

$$\forall\, \lambda \in \mathbb{R} \qquad \{z_0^{**} + \lambda \cdot (g_1, g_2) + \mu(-g_2, g_1) : \mu > 0\} \cap \mathcal{E}_{12} = \emptyset. \tag{40}$$

(38)-(40) is a complete characterization of $\mathcal{E}_{12}$ as a closed region between two parallel lines with (g_1, g_2) describing the direction of the line.

Let us first show (38). Because $\mathcal{E}_{12}$ does not have an extreme point whose coordinate in some dimension is a global extremum of $\mathcal{E}_{12}$ in that dimension, for any point $z_0 \in \mathcal{E}_{12}$ there are three unit length directions – for now let us denote them as $(e_1^{(i)}(z_0), e_2^{(i)}(z_0))$, $i = 1, 2, 3$, – such that $e_1^{(1)}(z_0)e_1^{(2)}(z_0) < 0$ and $e_2^{(1)}(z_0)e_2^{(3)}(z_0) < 0$ and $z_0 + \lambda(e_1^{(i)}(z_0), e_2^{(i)}(z_0)) \in \mathcal{E}_{12}$, $i = 1, 2, 3$ for any $\lambda \geq 0$.

It has to hold that two of these vectors are facing in the direction of opposite quadrants meaning their first coordinates have different signs and their second coordinates have different signs. Without a loss of generality, suppose these are vectors $(e_1^{(1)}(z_0), e_2^{(1)}(z_0))$ and $(e_1^{(2)}(z_0), e_2^{(2)}(z_0))$. If $(e_1^{(1)}(z_0), e_2^{(1)}(z_0)) \neq -(e_1^{(2)}(z_0), e_2^{(2)}(z_0))$, then by using the convexity of $\mathcal{E}_{12}$ we will be able to find convex combinations of $z_0 + \lambda(e_1^{(1)}(z_0), e_2^{(1)}(z_0))$ and $z_0 + \tilde{\lambda}(e_1^{(2)}(z_0), e_2^{(2)}(z_0))$, $\lambda, \tilde{\lambda} \geq 0$ that will belong to $\mathcal{E}_{12}$ and will form a quadrant $z_0 + \mathcal{O}_{s_1\, s_2}$ for some $s_1, s_2 \in \{+, -\}$, which will contradict the supposition that $\mathcal{E}_{12}$ does not contain any quadrants. Indeed, s_1 will be the sign of $e_1^{(1)}(z_0)$ if $|e_1^{(1)}(z_0)| > |e_1^{(2)}(z_0)|$ or the sign of $e_1^{(2)}(z_0)$ if $|e_1^{(2)}(z_0)| > |e_1^{(1)}(z_0)|$ (note that due to the suppositions in this case we cannot have $|e_1^{(1)}(z_0)| = |e_1^{(2)}(z_0)|$). Analogously, s_2 will be the sign of $e_2^{(1)}(z_0)$ if $|e_2^{(1)}(z_0)| > |e_2^{(2)}(z_0)|$ or the sign of $e_2^{(2)}(z_0)$ if $|e_2^{(2)}(z_0)| > |e_2^{(1)}(z_0)|$ (note that due to the suppositions in this case we cannot have $|e_2^{(1)}(z_0)| = |e_2^{(2)}(z_0)|$). Thus, it has to be that $(e_1^{(1)}(z_0), e_2^{(1)}(z_0)) = -(e_1^{(2)}(z_0), e_2^{(2)}(z_0))$. We also conclude that $(e_1^{(3)}(z_0), e_2^{(3)}(z_0))$ coincides with one of $(e_1^{(i)}(z_0), e_2^{(i)}(z_0))$, $i = 1, 2$, because if does not, then by using the convexity of $\mathcal{E}_{12}$ we will be able to show that $\mathcal{E}_{12}$ contains the quadrant $z_0 + \mathcal{O}_{s_1\, s_2}$ with s_1 being the sign of $e_1^{(3)}(z_0)$ and s_2 being the sign of $e_2^{(3)}(z_0)$, which is a contradiction. Thus, $(e_1^{(3)}(z_0), e_2^{(3)}(z_0))$ coincides with one of $(e_1^{(i)}(z_0), e_2^{(i)}(z_0))$, $i = 1, 2$. Without a loss of generality, we will take $e_1^{(1)}(z_0) > 0$ (we can swap $(e_1^{(1)}(z_0), e_2^{(1)}(z_0))$ and $(e_1^{(2)}(z_0), e_2^{(2)}(z_0))$ to achieve that if necessary).

Let us now show that for any $z_0 \in \mathcal{E}_{12}$ the direction $(e_1(z_0), e_2(z_0))$ is the same. Suppose that for two $z_0, \tilde{z}_0 \in \mathcal{E}_{12}$, $z_0 \neq \tilde{z}_0$, we have two different unit length vectors $(e_1(z_0), e_2(z_0))$ and $(e_1(\tilde{z}_0), e_2(\tilde{z}_0))$. Without a loss of generality, $e_1(z_0) \neq e_1(\tilde{z}_0)$. By convexity, $\mathcal{E}_{12}$ will contain all convex combinations of $z_0 + \lambda(e_1(z_0), e_2(z_0))$ and $\tilde{z}_0 + \tilde{\lambda}(e_1(\tilde{z}_0), e_2(\tilde{z}_0))$ for any $\lambda, \tilde{\lambda} \in \mathbb{R}$. Because $(e_1(z_0), e_2(z_0)) \neq (e_1(\tilde{z}_0), e_2(\tilde{z}_0))$, these convex combinations will give the whole $\mathbb{R}^2$, which contradicts the supposition that $\mathcal{E}_{12}$ does not contain any quadrants. Thus, all $(e_1(z_0), e_2(z_0))$ are the same and we can denote this direction as (g_1, g_2).

Now that we have established (38), we note that (39) and (40) just say that there are two straight lines in $\mathcal{E}_{12}$ (of course, with the direction of (g_1, g_2)) that form the boundary of $\mathcal{E}_{12}$. If, for example, (39) were violated, then $\mathcal{E}_{12}$ would have contained some quadrant $z_0 + \mathcal{O}_{s_1 s_2}$, where s_1 is the sign of g_2 and s_2 is the sign of $-g_1$, which would be a contradiction. Analogously with (40).

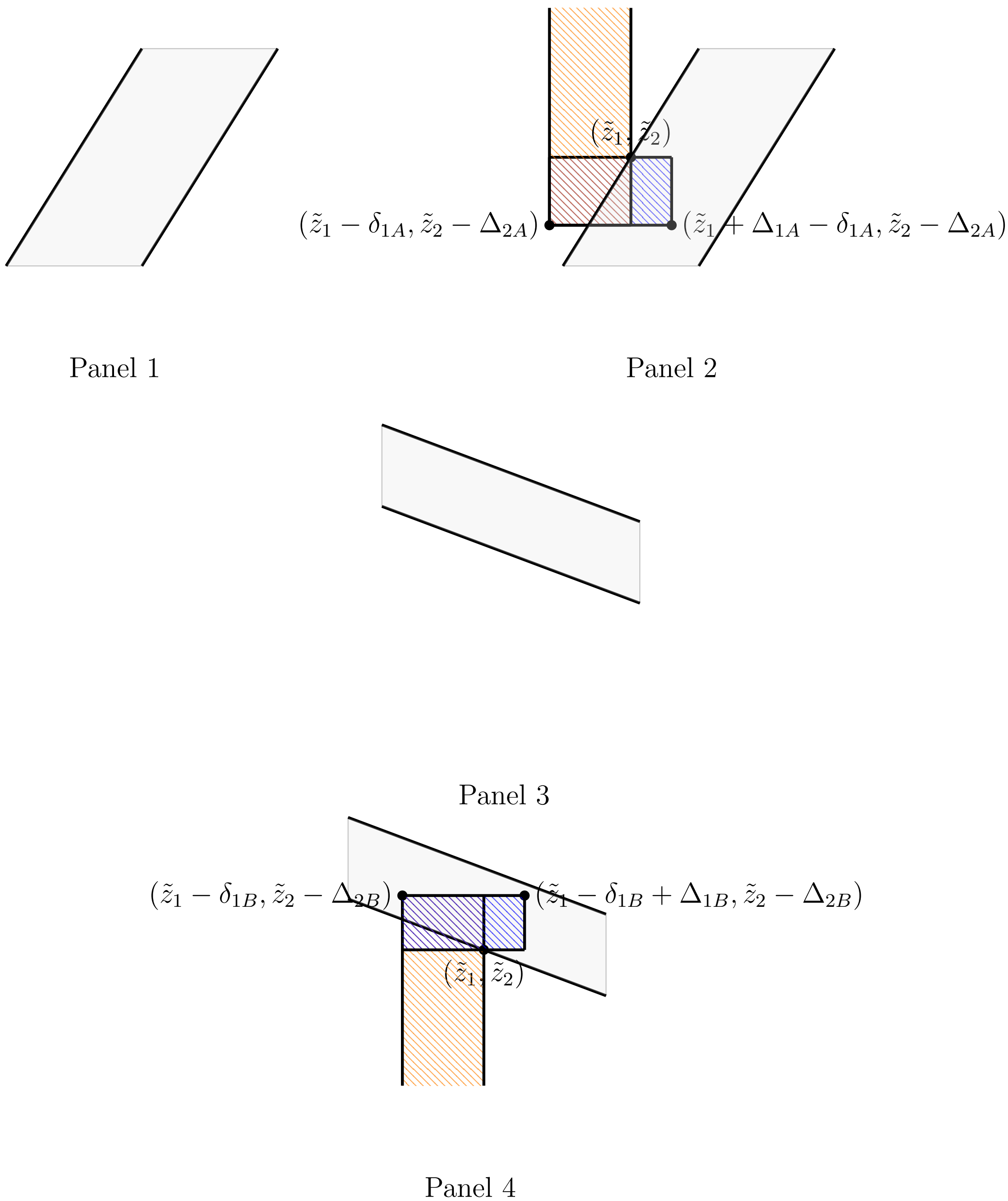
We can continue to suppose without a loss of generality that $g_1 > 0$. We now consider two situations.

The first situation is when $g_2 > 0$ (more generally can be described as g_1 and g_2 having the same sign). This situation is illustrated in in Panel 1 in Figure 21. Panel 2 in Figure 21 illustrates how in this case one can use the identification approach from the bottom left "corner" to show that there cannot be two different sets of parameters $(\Delta_{1A}, \Delta_{2A})$ and $(\delta_{1A}, \delta_{2A})$ that give the same function $Q_{1,2}(\cdot, \cdot)$ (defined earlier) everywhere. Indeed, suppose that there are such two sets of parameters and, without a loss of generality, $\Delta_{1A} > \delta_{1A}$ (then, necessarily, $\Delta_{2A} < \delta_{2A}$). Choose a point $(\tilde{z}_1, \tilde{z}_2)$ on the border of $\mathcal{E}_{12}$ such that the region $[\tilde{z}_1 - \delta_{1A}, \tilde{z}_1) \times [\tilde{z}_2, +\infty]$ is fully outside of $\mathcal{E}_{12}$ while the region $(\tilde{z}_1, \tilde{z}_1 + \Delta_{1A} - \delta_{1A}) \times (\tilde{z}_2 - \Delta_{2A}, \tilde{z}_{2A})$ has a non-empty intersection with the interior of $\mathcal{E}_{12}$. Taking $z_1^* = \tilde{z}_1 - \delta_{1A}$ and $z_2^* = \tilde{z}_2 - \Delta_{2A}$, we have that in this case the value of $Q_{1,2}(z_1^*, z_2^*)$ calculate when $(\Delta_{1A}, \Delta_{2A})$ is used is different from the value of $Q_{1,2}(z_1^*, z_2^*)$ calculate when $(\delta_{1A}, \delta_{2A})$ is used, thus giving us a contradiction. Note that the contradiction will be obtained with a positive probability since the difference in $Q_{1,2}(z_1, z_2)$ remains when using two different sets of parameters and using similar constructions can be made for (z_1, z_2) which are in $\mathcal{E}_{12}$ and in the neighborhood of such boundary point $(\tilde{z}_1, \tilde{z}_2)$.

The second situation is when $g_2 < 0$ (more generally can be described as g_1 and g_2 having different signs). This situation is illustrated in Panel 3 in Figure 21. Panel 4 in Figure 21 illustrates how in this case one can use the identification approach from the bottom left "corner" to show that there cannot be two different sets of parameters $(\Delta_{1B}, \Delta_{2B})$ and $(\delta_{1B}, \delta_{2B})$ that give the same function $Q_{1,\overline{2}}(\cdot, \cdot)$ (defined earlier) everywhere. Recall from earlier that $\Delta_{1B}, \delta_{1B} > 0$ and $\Delta_{2B}, \delta_{2B} < 0$. Without a los sf generality, we can take $\delta_{1B} < \Delta_{1B}$ (then, necessarily, we must have $\delta_{2B} > \Delta_{2B}$). We can choose a point $(\tilde{z}_1, \tilde{z}_2)$ on the border of $\mathcal{E}_{12}$ such that the region $[\tilde{z}_1 - \delta_{1B}, \tilde{z}_1) \times (-\infty, \tilde{z}_2]$ is fully outside of $\mathcal{E}_{12}$ while the region $(\tilde{z}_1, \tilde{z}_1 + \Delta_{1B} - \delta_{1B}) \times (\tilde{z}_2, \tilde{z}_2 - \Delta_{2B})$ has a non-empty intersection with the interior of $\mathcal{E}_{12}$. Taking $z_1^* = \tilde{z}_1 - \delta_{1B}$ and $z_2^* = \tilde{z}_2 - \Delta_{2B}$, we have that in this case the value of $Q_{1,\overline{2}}(z_1^*, z_2^*)$ calculate when $(\Delta_{1B}, \Delta_{2B})$ is used is different from the value of $Q_{1,\overline{2}}(z_1^*, z_2^*)$ calculate when $(\delta_{1B}, \delta_{2B})$ is used, thus giving us a contradiction. Note that the contradiction will be obtained with a positive probability since the difference in $Q_{1,\overline{2}}(z_1, z_2)$ remains when using two different sets of parameters and using similar constructions can be made for (z_1, z_2) which are in $\mathcal{E}_{12}$ and in the neighborhood of such boundary point $(\tilde{z}_1, \tilde{z}_2)$.

Thus, we showed that in every scenario the pair (Δ_1, Δ_2) is identified from the observed choice probabilities as in (31).

FIGURE 21: Fourth illustration of identification in Step 4.



Having established that (Δ_1, Δ_2) is identified from (31), we can go back to our thresholds problem and notations and conclude that thresholds $\alpha^{(1)}_{2,2,1,\ldots,1}$ or $\alpha^{(2)}_{2,2,1,\ldots,1}$ are identified. What we showed is that once we know $\alpha^{(1)}_{j_1,j_2+1,1,\ldots,1}$ and $\alpha^{(2)}_{j_1+1,j_2,1,\ldots,1}$, then we can identify $\alpha^{(1)}_{j_1+1,j_2+1,1,\ldots,1}$ and $\alpha^{(2)}_{j_1+1,j_2+1,1,\ldots,1}$. Applying this sequentially, we can show that any $\alpha^{(1)}_{j_1,2,1,\ldots,1}$ and $\alpha^{(2)}_{j_1,2,1,\ldots,1}$ as well as any $\alpha^{(1)}_{2,j_2,1,\ldots,1}$ and $\alpha^{(2)}_{2,j_2,1,\ldots,1}$ are identified. Then we will apply the same result to show that any $\alpha^{(1)}_{3,j_2,1,\ldots,1}$ and $\alpha^{(2)}_{3,j_2,1,\ldots,1}$ are identified, and so on. In this way, we will show that any $\alpha^{(1)}_{j_1,j_2,1,\ldots,1}$ and $\alpha^{(2)}_{j_1,j_2,1,\ldots,1}$ are identified.

In our example on the left panel in Figures 15-17, we can now identify all the thresholds. Since we in general we will have more than 2 dimensions for the response variable, then we need to discuss identification of thresholds when we vary indices in other dimensions as well. This is done in Step 5.

Step 5. It is enough for us to describe how to identify all the thresholds when we vary indices in three dimensions – without a loss of generality, we can take $\alpha^{(1)}_{j_1,j_2,j_3,1,\ldots,1}$, $\alpha^{(2)}_{j_1,j_2,j_3,1,\ldots,1}$ and $\alpha^{(3)}_{j_1,j_2,j_3,1,\ldots,1}$, – as the extension to other dimensions will be analogous.

Just like in Step 4, it is enough for us to establish the identification of thresholds $\alpha^{(h)}_{j_1,j_2,j_3,1,\ldots,1}$, $h = 1, 2, 3$, where $j_\ell \in \{2, M_\ell - 1\}$, $\ell = 1, 2, 3$, as from Step 4 we know already that the thresholds

- $\alpha^{(1)}_{j_1,j_2,j_3,1,\ldots,1}$, where $j_1 \in \{1, M_1\}$, $j_2 \in \{2, M_2 - 1\}$, $j_3 \in \{2, M_3 - 1\}$,
- $\alpha^{(2)}_{j_1,j_2,j_3,1,\ldots,1}$, where $j_2 \in \{1, M_2\}$, $j_1 \in \{2, M_1 - 1\}$, $j_3 \in \{2, M_3 - 1\}$,
- $\alpha^{(3)}_{j_1,j_2,j_3,1,\ldots,1}$, where $j_3 \in \{1, M_3\}$, $j_1 \in \{2, M_1 - 1\}$, $j_2 \in \{2, M_2 - 1\}$,

are identified. Just like in Step 4, we may have situations when for fixed (j_1, j_2, j_3), where $j_\ell \in \{2, M_\ell - 1\}$, $\ell = 1, 2, 3$, one of two of the three thresholds parameters $\alpha^{(h)}_{j_1,j_2,j_3,1,\ldots,1}$, $h = 1, 2, 3$, are known. However, we need an identification strategy when all three threshold parameters $\alpha^{(h)}_{2,2,2,1,\ldots,1}$, $h = 1, 2, 3$, are unknown.

Just like in Step 4, we can start identification from different "corners". These "corners" are now in three dimensions and are harder to label with words like we did before when used "bottom left corner" or "top left corner". However, we can now describe identification stemming from these different three-dimensional "corners" as identification happening in the direction of the orthant $\mathcal{O}_{s_1\,s_2\,s_3}$, where $s_d \in \{+, -\}$, $d = 1, 2, 3$, where

$$\mathcal{O}_{s_1\,s_2\,s_3} = \{(s_1\lambda_1, s_2\lambda_2, s_3\lambda_3) : \lambda_d \geq 0, d = 1, 2, 3\}.$$

If the identification proceeds in the direction of $\mathcal{O}_{+++}$, we first try to establish the identification of thresholds $\alpha^{(h)}_{2,2,2,1,\ldots,1}$, $h = 1, 2, 3$, and from Step 4 we know already that the thresholds $\alpha^{(1)}_{1,2,2,1,\ldots,1}$,

$\alpha^{(2)}_{2,1,2,1,\ldots,1}$ and $\alpha^{(3)}_{2,2,1,1,\ldots,1}$ are identified. For that we consider the following observed probability $P\left(\bigcap_{h=1}^{3}\left(Y^{c_h} = y_2^{(h)}\right), \bigcap_{d>3}\left(Y^{c_d} = y_1^{(d)}\right) \mid x\right)$. Taking $x_{d,1} \to -\infty$ for all $d > 3$, in the limit we identify

$$P\left(\bigcap_{h=1}^{3}\left(Y^{c_h} = y_2^{(h)}\right) \mid x\right) = \sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{\ell_3=0}^{1}(-1)^{\ell_1+\ell_2+\ell_3+1}\cdot$$
$$F_{1,2,3}\left(\ell_1\alpha^{(1)}_{2,2,2,1,\ldots,1} + (1-\ell_1)\alpha^{(1)}_{1,2,2,1,\ldots,1} - x_1\beta_1,\right.$$
$$\left.\ell_2\alpha^{(2)}_{2,2,2,1,\ldots,1} + (1-\ell_2)\alpha^{(1)}_{2,1,2,1,\ldots,1} - x_2\beta_2, \ell_3\alpha^{(3)}_{2,2,2,1,\ldots,1} + (1-\ell_3)\alpha^{(3)}_{2,2,1,1,\ldots,1} - x_3\beta_3\right) \quad (41)$$

where $F_{1,2,3}$ denotes the joint c.d.f. of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$, which is known by Theorem 5. The question is whether it is possible to recover $\alpha^{(h)}_{2,2,2,1,\ldots,1}$, $h = 1, 2, 3$ from the observed probabilities $P\left(\cap_{h=1}^{3}\left(Y^{c_h} = y_2^{(h)}\right) \mid x\right)$.

Analogously to Step 4 and (31), we can be reformulate this problem as the problem of showing that there is only one set of parameters $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$, $\Delta_{iA} > 0$, $i = 1, 2, 3$, such that for any (z_1, z_2, z_3)

$$Q_{1,2,3}(z_1, z_2, z_3) = \sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{\ell_d=0}^{1}(-1)^{\ell_1+\ell_2+\ell_3+1} F_{123}\Big(\ell_1\Delta_1 + z_1, \ell_2\Delta_2 + z_2, \ell_3\Delta_3 + z_3\Big), \quad (42)$$

where $Q_{1,2,3}(z_1, z_2, z_3)$ is known and, of course, denotes the probability of choice. Vector (z_1, z_2, z_3) can take any value in $\mathbb{R}^3$.

If the identification proceeds, for instance, in the direction of the orthant $\mathcal{O}_{+-+}$, then analogously the identification problem can be reformulated as the problem of showing that there is only one set of parameters $(\Delta_{1B}, \Delta_{2B}, \Delta_{3B})$, $\Delta_{1B}, \Delta_{3B} > 0$, $\Delta_{2B} < 0$, such that for any (z_1, z_2, z_3)

$$Q_{1,\bar{2},3}(z_1, z_2, z_3) = \sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{\ell_d=0}^{1}(-1)^{\ell_1+\ell_2+\ell_3+1} F_{1,\bar{2},\bar{3}}\Big(\ell_1\Delta_{1B} + z_1, \ell_2\Delta_{2B} + z_2, \ell_3\Delta_{3B} + z_3\Big), \quad (43)$$

where $Q_{1,\bar{2},3}(z_1, z_2, z_3)$ is known and, of course, denotes the probability of choice. Vector (z_1, z_2, z_3) can take any value in $\mathbb{R}^3$.

If the identification proceeds, for instance, in the direction of the orthant $\mathcal{O}_{-++}$, the analogously the identification problem can be reformulated as the problem of showing that there is only one set of parameters $(\Delta_{1C}, \Delta_{2C}, \Delta_{3C})$, $\Delta_{1C} < 0$ and $\Delta_{2C}, \Delta_{3C} > 0$, such that for any (z_1, z_2, z_3)

$$Q_{1,2,\bar{3}}(z_1, z_2, z_3) = \sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{\ell_d=0}^{1}(-1)^{\ell_1+\ell_2+\ell_3+1} F_{1,2,\bar{3}}\Big(\ell_1\Delta_{1C} + z_1, \ell_2\Delta_{2C} + z_2, \ell_3\Delta_{3C} + z_3\Big), \quad (44)$$

where $Q_{\bar{1},2,3}(z_1, z_2, z_3)$ is known and, of course, denotes the probability of choice. Vector (z_1, z_2, z_3) can take any value in $\mathbb{R}^3$.

Denote the support of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ as $\mathcal{E}_{123}$.

(i) The first situation we consider is when $\mathcal{E}_{123}$ has an extreme point and at least one of the coordinates of this extreme point is the global maximum or minimum of $\mathcal{E}_{123}$ in that dimension. In this case, the uniqueness of thresholds can be proven using identification in the direction of any orthant $\mathcal{O}_{s_1\, s_2\, s_3}$. We will use identification in the direction of the orthant $\mathcal{O}_{+++}$. In this case, the proof of uniqueness should be based n the properties of the function $Q_{1,2,3}$.

Suppose that there is another set of parameters $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ with $\delta_{iA} > 0$ for $i = 1, 2, 3$, such that for any (z_1, z_2, z_3)

$$Q_{1,2,3}\,(z_1, z_2, z_3) = \sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{\ell_d=0}^{1} (-1)^{\ell_1+\ell_2+\ell_3+1}\, F_{123}\Big(\ell_1\delta_{1A} + z_1, \ell_2\delta_{2A} + z_2, \ell_3\delta_{3A} + z_3\Big). \tag{45}$$

The component-wise monotonicity in each δ_{1A}, $i = 1, 2, 3$, of the right-hand side of (45) implies that for these two different sets of parameters to give the same choice probabilities, one should have $\delta_{hA} < \Delta_{hA}$ for at least one (and at most two) $h = 1, 2, 3$. We can suppose, without a loss of generality, that $\delta_{1A} < \Delta_{1A}$ and $\delta_{2A} > \Delta_{2A}$ (the relation between δ_{3A} and Δ_{3A} does not matter).[27]

Denote

$$R_{\Delta_A}(z_1, z_2, z_3) = \bigtimes_{h=1}^{3}[z_h, z_h + \Delta_{hA}], \quad R_{\delta_A}(z_1, z_2, z_3) = \bigtimes_{h=1}^{3}[z_h, z_h + \delta_{hA}].$$

The technicalities of establishing that the vectors $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ and $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ coincide, are analogous to Step 4. We establish that we can always choose (z_1, z_2, z_3) such that one of the regions $R_{\Delta_A}(z_1, z_2, z_3)\backslash R_{\delta_A}(z_1, z_2, z_3)$ and $R_{\delta_A}(z_1, z_2, z_3)\backslash R_{\Delta_A}(z_1, z_2, z_3)$ is outside of the interior of $\mathcal{E}_{123}$, whereas the other one has an intersection with the interior of $\mathcal{E}_{123}$. If we establish this fact, then we obtain a contradiction with the fact that $R_{\Delta_A}(z_1, z_2, z_3)$ and $R_{\delta_A}(z_1, z_2, z_3)$ have the same probability mass.

Indeed, denote an extreme point described in the condition of this case as $z_0 = (z_{10}, z_{20}, z_{30})$. Suppose, for instance, that z_{30} is the global minimum of $\mathcal{E}_{123}$ in the third dimension. Suppose for now that neither of other coordinates z_{10} and z_{20} is a global minimum of $\mathcal{E}_{123}$ in the respective dimension. Choose $(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30})$ such that

$$\tilde{z}_{30} + \min\{\delta_{3A}, \Delta_{3A}\} = z_{30}, \quad \tilde{z}_{20} + \Delta_{2A} = z_{20}, \quad \tilde{z}_{10} + \delta_{1A} = z_{10} \tag{46}$$

(recall that $\delta_{1A} < \Delta_{1A}$ and $\delta_{2A} > \Delta_{2A}$). Our discussion below uses the fact that due to assumptions on the boundary of the distribution of ε, the singleton $\{z_0\}$ has a zero probability mass in the distribution

[27] In case of different relations, the subsequent role of dimensions would change.

of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$.

- Consider first $\Delta_{3A} < \delta_{3A}$. $R_\Delta(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30}) \backslash R_\delta(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30})$ does not intersect with the interior $\mathcal{E}_{123}$ (touches it at z_0) and, thus, has a zero probability mass, whereas the region $R_{\delta_A}(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30}) \backslash R_{\Delta_A}(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30})$ does have an intersection with the interior of $\mathcal{E}_{123}$ due to the supposition that neither of other coordinates z_{10} and z_{20} is a global minimum of $\mathcal{E}_{123}$ in the respective dimension, and this intersection has a positive probability mass.

- If $\delta_{3A} < \Delta_{3A}$, then the region

$$R_{\delta_A}(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30}) \backslash R_{\Delta_A}(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30})$$

  does not intersect with the interior of $\mathcal{E}_{123}$ and, thus, has a zero probability mass, whereas the region $R_{\Delta_A}(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30}) \backslash R_{\delta_A}(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30})$ does have an intersection with the interior of $\mathcal{E}_{123}$ due to the supposition that neither of other coordinates z_{10} and z_{20} is a global minimum of $\mathcal{E}_{123}$ in the respective dimension, and this intersection has a positive probability mass.

Suppose now that at least one other coordinate – z_{10} or z_{20} – is the global minimum of $\mathcal{E}_{123}$ in the respective dimension. Then it is possible that both regions $R_{\Delta_A}(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30}) \backslash R_{\delta_A}(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30})$ and $R_{\delta_A}(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30}) \backslash R_\Delta(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30})$ lie outside of the interior of $\mathcal{E}_{123}$ and, thus, have zero probability mass if $(\tilde{z}_{10}, \tilde{z}_{20}, \tilde{z}_{30})$ is chosen as in (46). In this case instead of the extreme point $z_0 = (z_{10}, z_{20}, z_{30})$ described above we consider $z_{*1} = (z_{11}, z_{21}, z_{31})$ which is in a small neighborhood of z_0 and is in the interior of $\mathcal{E}_{123}$. We then move z_{*1} in the direction $(0, 0, -1)$ until we hit the boundary. We denote the final point as $z_{*2} = (z_{12}, z_{22}, z_{32})$. Choosing z_{*2} in such a way guarantees that in each of the four orthants $z_{*2} + \mathcal{O}_{s_1 s_2 +}$, $s_1, s_2 \in \{+, -\}$, a rectangle $(z_{12} + s_1 a_1) \times (z_{22} + s_2 a_2) \times (z_{32} + a_3)$ has a non-empty intersection with the interior of $\mathcal{E}_{123}$ for any $a_1 > 0$, $a_2 > 0$, $a_3 > 0$.

Choose $(\tilde{z}_{12}, \tilde{z}_{22}, \tilde{z}_{32})$ such that

$$\tilde{z}_{32} + \min\{\delta_{3A}, \Delta_{3A}\} = z_{32}, \quad \tilde{z}_{22} + \Delta_{2A} = z_{22}, \quad \tilde{z}_{12} + \delta_{1A} = z_{12}$$

and end up with the following two situations:

- if $\Delta_{3A} < \delta_3$, then the region

$$R_\Delta(\tilde{z}_{12}, \tilde{z}_{22}, \tilde{z}_{32}) \backslash R_{\delta_A}(\tilde{z}_{12}, \tilde{z}_{20}, \tilde{z}_{30})$$

  does not intersect with the interior of $\mathcal{E}_{123}$ and, thus, has a zero probability mass, whereas the region $R_{\delta_A}(\tilde{z}_{12}, \tilde{z}_{22}, \tilde{z}_{32}) \backslash R_{\Delta_A}(\tilde{z}_{12}, \tilde{z}_{22}, \tilde{z}_{32})$ does have an intersection with the interior of $\mathcal{E}_{123}$ due to the property of orthants described above, and this intersection has a positive probability mass;

- If $\delta_{3A} < \Delta_{3A}$, then the region

$$R_{\delta_A}(\tilde{z}_{12}, \tilde{z}_{22}, \tilde{z}_{32}) \backslash R_{\Delta_A}(\tilde{z}_{12}, \tilde{z}_{22}, \tilde{z}_{32})$$

  does not intersect with the interior of $\mathcal{E}_{123}$ and, thus, has a zero probability mass, whereas the region $R_{\Delta_A}(\tilde{z}_{12}, \tilde{z}_{22}, \tilde{z}_{32}) \backslash R_{\delta_A}(\tilde{z}_{12}, \tilde{z}_{22}, \tilde{z}_{32})$ has an intersection with the interior of $\mathcal{E}_{123}$ due to the property of orthants described above, and this intersection has a strictly positive probability mass.

Thus, in this case we are able to obtain a contradiction that both (44) and (45) can hold simultaneously if $\Delta_A \neq \delta_A$.

Note that contradictions in this case can be obtained with a positive probability since a discontinuity in the probability of two differenced regions will also be obtained for points $z \in \mathcal{E}_{12}$ in the neighborhood of z_0 and also for points $z \in \mathcal{E}_{12}$ in the neighborhood of z_{2*}.

Analogous constructions and contradictions can be obtained in the case when an extreme point is a global maximum in the third dimension or when it is a global optimum in the first or second dimensions.

(ii) Our second case is when $\mathcal{E}_{123}$ contains a whole orthant $z^* + \mathcal{O}_{s_1\, s_2\, s_3}$ for some $z^* \in \mathcal{E}_{123}$ and some (s_1, s_2, s_3), $s_d \in \{+, -\}$, $d = 1, 2, 3$. Then the proof of the uniqueness thresholds will proceed in a way analogous to the similar second case in Step 4. The contradictions will be obtained when employing identification in the direction of the orthant $\mathcal{O}_{s_1\, s_2\, s_3}$.

Suppose, e.g., that $s_d = +$ for all $d = 1, 2, 3$. Then we use the function $Q_{1,2,3}$ and note that for *just one value* of (z_1^*, z_2^*, z_3^*), the relation (44) gives a surface of $(\Delta_1, \Delta_2, \Delta_3)$ that satisfies that relation. That surface can be expressed by a function $\Delta_1(\Delta_2, \Delta_3)$ which is strictly decreasing coordinatewise in Δ_2 and Δ_3. On this surface there is a minimum value that Δ_1 can take and that we can denote as $\tilde{\underline{\Delta}}(z_1^*, z_2^*, z_3^*)$. This minimum value uniquely solves the following equation:

$$Q_{1,2,3}\,(z_1^*, z_2^*, z_3^*) = \sum_{\ell_1=0}^{1} \sum_{\ell_2=0}^{1} \sum_{\ell_d=0}^{1} (-1)^{\ell_1+\ell_2+\ell_3+1}\, F_{1,2,3}\Big(\ell_1 \tilde{\underline{\Delta}}(z_1^*, z_2^*, z_3^*) + z_1^*, \ell_2 \cdot (+\infty) + z_2^*, \ell_3 \cdot (+\infty) + z_3^*\Big),$$

where we take $\ell_d \cdot (+\infty) = 0$ when $\ell_d = 0$, $d = 1, 2$. Analogously to Step 4, we can obtain a contradiction by showing that because of $\delta_{1A} < \Delta_{1A}$, we can fix $z_{1,\lambda_1}^* = z_1^* + \lambda_1$, $\lambda_1 \geq 0$ and choose $\lambda_2, \lambda_3 \geq 0$ large enough so that $z^*_{2,\lambda_2} = z^*_2 + \lambda_2$ and $z^*_{3,\lambda_3} = z^*_3 + \lambda_3$ to be large enough so that $\underline{\tilde{\Delta}}(z^*_{1,\lambda_1}, z^*_{2,\lambda_2}, z^*_{3,\lambda_3}) > \delta_{1A}$.[28] Analogously to Step 4, this would be shown from the fact that by making $\lambda_2, \lambda_3 \geq 0$ large enough, we will have that

$$\frac{\sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{\ell_d=0}^{1}(-1)^{\ell_1+\ell_2+\ell_3+1}F_{1,2,3}\Big(\ell_1\Delta_1 + z^*_{1,\lambda_1}, \ell_2\Delta_2 + z^*_{2,\lambda_2}, \ell_3\Delta_3 + z^*_{3,\lambda_3}\Big)}{\sum_{\ell_1=0}^{1}\sum_{\ell_2=0}^{1}\sum_{\ell_d=0}^{1}(-1)^{\ell_1+\ell_2+\ell_3+1}F_{1,2,3}\Big(\ell_1\delta_{1A} + z^*_{1,\lambda_1}, \ell_2\cdot(+\infty) + z^*_{2,\lambda_2}, \ell_3\cdot(+\infty) + z^*_{3,\lambda_3}\Big)} > 1. \tag{47}$$

The fact that we can choose $(z^*_{1,\lambda_1}, z^*_{2,\lambda_2}, z^*_{3,\lambda_3})$ such that (47) holds follows from the properties of the multivariate c.d.f. and can be especially easily seen when ε_1, ε_2 and ε_3 are mutually independent as then (47) can be rewritten as

$$\frac{\prod_{h=1}^{3}\left(F_h\left(\Delta_{hA} + z^*_{h,\lambda_h}\right) - F_h\left(z^*_{h,\lambda_h}\right)\right)}{\left(F_1\left(\delta_1 + z^*_{1,\lambda_1}\right) - F_1\left(z^*_{1,\lambda_1}\right)\right)\cdot\left(1 - F_2\left(z^*_{2,\lambda_2}\right)\right)\cdot\left(1 - F_3\left(z^*_{3,\lambda_3}\right)\right)},$$

as it is obvious that we can fix z^*_{1,λ_1} and take $\lambda_2, \lambda_3 \to \infty$, in which case we have

$$\frac{F_1\left(\Delta_{1A} + z^*_{1,\lambda_1}\right) - F_1\left(z^*_{1,\lambda_1}\right)}{F_1\left(\delta_{1A} + z^*_{1,\lambda_1}\right) - F_1\left(z^*_{1,\lambda_1}\right)} > 1$$

and

$$\frac{F_2\left(\Delta_{2A} + z^*_{2,\lambda_2}\right) - F_2\left(z^*_{2,\lambda_2}\right)}{1 - F_2\left(z^*_{2,\lambda_2}\right)} \to 1, \frac{F_3\left(\Delta_{3A} + z^*_{3,\lambda_3}\right) - F_3\left(z^*_{3,\lambda_3}\right)}{1 - F_3\left(z^*_{3,\lambda_3}\right)} \to 1.$$

For a general c.d.f. $F_{1,2,3}$, the property (47) can be shown by employing properties of the copula $C(\cdot,\cdot,\cdot)$ that corresponds to $F_{1,2,3}$ – namely, that $|C(u,v,w) - u| \to 0$ as $v \to 1$ and $w \to 1$.

In summary, by properties of $\mathcal{E}_{12}$ (convexity and non-empty interior) there will be a positive measure of $z^* \in \mathcal{E}_{123}$ such that $z^* + \mathcal{O}_{+++}$ is contained in $\mathcal{E}_{123}$. Therefore, for a positive measure of $(z^*_{1,\lambda_1}, z^*_{2,\lambda_2}, z^*_{3,\lambda_3})$ we obtain that $\underline{\tilde{\Delta}_1}(z^*_{1,\lambda_1}, z^*_{2,\lambda_2}, z^*_{3,\lambda_3}) > \delta_{1A}$ giving us the contradiction that parameter δ_{1A} together with δ_{2A}, δ_{3A} is observationally equivalent to Δ_{1A} and Δ_{2A}, Δ_{3A}, where $\Delta_{1A} > \delta_{1A}$. This contradiction allows us to conclude that $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ is identified from (31). Note that we can modify the proof of this subcase by instead considering a fixed λ_2 or λ_3 and taking thhe other two lambdas to $+\infty$.

If, for instance, $s_1 = +$ and $s_2 = s_3 = -$, then we use $Q_{1,\overline{2},\overline{3}}$ defined in terms of $F_{1,\overline{2},\overline{3}}$ to obtain the uniqueness of thresholds. In an analogous way we would approach the case of any (s_1, s_2, s_3).

[28]If we had $\Delta_{1A} < \delta_{1A}$, we would reverse the role of these two values and instead we would consider the surface of $(\delta_1, \delta_2, \delta_3)$ defined by relation (45).

(iii) Finally, we consider the intermediate case when (a) $\mathcal{E}_{123}$ does not have an extreme point whose coordinate in some dimension is a global extremum of $\mathcal{E}_{123}$ in that dimension, and at the same time (b) $\mathcal{E}_{123}$ does not contain any orthants in the form $z^* + \mathcal{O}_{s_1\, s_2\, s_3}$.

In this case take some point (z_{10}, z_{20}, z_{30}) in the interior of $\mathcal{E}_{123}$. Then at least one of the three half-lines $(z_{10} + \lambda_1, z_{20}, z_{30})$, $\lambda_1 > 0$, $(z_{10}, z_{20} + \lambda_2, z_{30})$, $\lambda_2 > 0$, and $(z_{10}, z_{20}, z_{30} + \lambda_3)$ is not fully contained in $\mathcal{E}_{123}$. (Indeed, it all three half-lines were fully in $\mathcal{E}_{123}$, then convexity of $\mathcal{E}_{123}$ would imply that the whole region $(z_{10} + \lambda_1, z_{20} + \lambda_2, z_{30} + \lambda_3)$ is in $\mathcal{E}_{123}$, which contradicts the characterization of this case). Without a loss of generality, suppose that it is $(z_{10}, z_{20}, z_{30}) + \lambda_3$ that is not fully contained in $\mathcal{E}_{123}$. Because of the convexity of $\mathcal{E}_{123}$, there is a point $(z_{10}, z_{20}, z_{30} + \overline{\lambda}_3)$ that belongs to $\mathcal{E}_{123}$ but no point $(z_{10}, z_{20}, z_{30} + \lambda_3)$, $\lambda_3 > \overline{\lambda}_3$, is in $\mathcal{E}_{123}$. Naturally, $(z_{10}, z_{20}, z_{30} + \overline{\lambda}_3)$ is on the border of $\mathcal{E}_{123}$. Denote $\tilde{z} = (\tilde{z}_1, \tilde{z}_2, \tilde{z}_3) = (z_{10}, z_{20}, z_{30} + \overline{\lambda}_3)$.

Our next observation is that at least one of the four orthants $\tilde{z} + \mathcal{O}_{s_1\, s_2\, +}$, $s_1, s_2 \in \{+, -\}$, does not contain any points from $\mathcal{E}_{123}$. (Indeed, if all of these orthants contained points from $\mathcal{E}_{123}$, we would be able to find a convex combination of these points which would have the first two coordinates exactly at $\tilde{z}_1$ and $\tilde{z}_2$, whereas the third covariate would be strictly greater than $\tilde{z}_3$, which would contradict the choice of $\tilde{z}_3$.) Let us call it an "empty orthant". At the same time, since by the characterization of this case $\tilde{z}_3$ cannot be a global maximum in the third dimension, then at least one of these four orthants has a non-empty intersection with the interior of $\mathcal{E}_{123}$ – more precisely, due to convexity of $\mathcal{E}_{123}$, a 3-dimensional rectangle $conv(\tilde{z}_1, \tilde{z}_1 + s_1 t_1) \times conv(\tilde{z}_2, \tilde{z}_2 + s_2 t_2) \times [\tilde{z}_3, \tilde{z}_3 + t_3]$ for any $t_d > 0$, $d = 1, 2, 3$, has a non-empty intersection with the interior of $\mathcal{E}_{123}$. Let us call it an "intersecting orthant". Here $conv(a_1, a_2)$ denotes a univariate interval connecting points a_1 and a_2. Since we only have four orthants $\mathcal{O}_{s_1\, s_2\, +}$, $s_d \in \{+, -\}$, $d = 1, 2$, then among them there will always be an "intersecting orthant" that is adjacent to an empty orthant among them. By adjacent we mean an orthant that has one and only one change of sign in first two dimension while maintaining the sign in the third dimension to be $+$. For instance, $\tilde{z} + \mathcal{O}_{+-+}$ is adjacent to both $\tilde{z} + \mathcal{O}_{+++}$ and $\tilde{z} + \mathcal{O}_{--+}$ but not to $\tilde{z} + \mathcal{O}_{-++}$. We also note that by construction of $\tilde{z}$, in any orthant $\tilde{z} + \mathcal{O}_{s_1\, s_2\, -}$, $s_1, s_2 \in \{+, -\}$, a 3-dimensional rectangle $conv(\tilde{z}_1, \tilde{z}_1 + s_1 t_1) \times conv(\tilde{z}_2, \tilde{z}_2 + s_2 t_2) \times [\tilde{z}_3, \tilde{z}_3 - t_3]$ for any $t_d > 0$, $d = 1, 2, 3$, has a non-empty intersection with the interior of $\mathcal{E}_{123}$. In other words,

(*Situation S1*) Suppose $\tilde{z} + \mathcal{O}_{--+}$ is an "empty orthant". Then the uniqueness of the thresholds can be proven by employing the identification in the direction of the orthant $\mathcal{O}_{+++}$. We will, thus, construct the proof by showing that there cannot be two sets of thresholds $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ and $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ that can give the same $Q_{1,2,3}$ everywhere.

First, suppose that $\tilde{z} + \mathcal{O}_{+-+}$ is an "intersecting orthant", which is adjacent to $\tilde{z} + \mathcal{O}_{--+}$. Consider[29] the following two 3-dimensional rectangles:

$$\mathcal{T}_1 = [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times [\tilde{z}_2 - \delta_{2A}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]$$
$$\mathcal{T}_2 = [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times [\tilde{z}_2 - \delta_{2A}, \tilde{z}_2 - \delta_{2A} + \Delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}].$$

If $\delta_{3A} < \Delta_{3A}$, then

$$\mathcal{T}_1 \backslash \mathcal{T}_2 = [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times (\tilde{z}_2 - \delta_{2A} + \Delta_{2A}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]$$

is in the closure of $\mathcal{O}_{--+}$ and, thus, has zero probability, whereas $\mathcal{T}_2 \backslash \mathcal{T}_1 = \mathcal{T}_{2a} \cup \mathcal{T}_{2b}$ with

$$\mathcal{T}_{2a} = (\tilde{z}_1, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times \times [\tilde{z}_2 - \delta_{2A}, \tilde{z}_2 - \delta_{2A} + \Delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}]$$
$$\mathcal{T}_{2b} = [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times [\tilde{z}_2 - \delta_{2A}, \tilde{z}_2 - \delta_{2A} + \Delta_{2A}] \times (\tilde{z}_3 + \delta_{3A}, \tilde{z}_3 + \Delta_{3A}].$$

Note that the rectangle $\mathcal{T}_{2a}$ is in $\tilde{z} + \mathcal{O}_{+-+}$, and has a strictly positive probability mass. This gives us a contradiction with the supposition that both $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ and $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ give the same observable $Q(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1 - \delta_{1A}$, $z_2 = \tilde{z}_2 - \delta_{2A}$ and $z_3 = \tilde{z}_3$. If $\delta_{3A} > \Delta_{3A}$, then $\mathcal{T}_1 \backslash \mathcal{T}_2 = \mathcal{T}_{1a} \cup \mathcal{T}_{1b}$, where

$$\mathcal{T}_{1a} = [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times (\tilde{z}_2 - \delta_{2A} + \Delta_{2A}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]$$
$$\mathcal{T}_{1b} = [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times [\tilde{z}_2 - \delta_{2A}, \tilde{z}_2] \times (\tilde{z}_3 + \Delta_{3A}, \tilde{z}_3 + \delta_{3A}].$$

Both $\mathcal{T}_{1a}$ and $\mathcal{T}_{1b}$ are in $\tilde{z} + \mathcal{O}_{--+}$ and, thus, have zero probability, whereas

$$\mathcal{T}_2 \backslash \mathcal{T}_1 = (\tilde{z}_1, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times \times [\tilde{z}_2 - \delta_{2A}, \tilde{z}_2 - \delta_{2A} + \Delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}]$$

is in $\tilde{z} + \mathcal{O}_{+-+}$ and has a non-zero probability. Once again, this gives us a contradiction with the supposition that both $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ and $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ give the same observable $Q(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1 - \delta_{1A}$, $z_2 = \tilde{z}_2 - \delta_{2A}$ and $z_3 = \tilde{z}_3$.

Second, suppose $\tilde{z} + \mathcal{O}_{-++}$ is an "intersecting orthant" with $\tilde{z} + \mathcal{O}_{--+}$ continuing to be an "empty orthant". Consider the following two 3-dimensional rectangles:

$$\mathcal{T}_1 = [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2 - \Delta_{2A} + \delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]$$
$$\mathcal{T}_2 = [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}].$$

[29] Recall that we supposed that $\delta_{1A} < \Delta_{1A}$ and $\delta_{2A} > \Delta_{2A}$.

If $\delta_{3A} < \Delta_{3A}$, then

$$\mathcal{T}_1 \backslash \mathcal{T}_2 = [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times (\tilde{z}_2, \tilde{z}_2 - \Delta_{2A} + \delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]$$

is in $\tilde{z} + \mathcal{O}_{-++}$, and since the latter is an "intersecting orthant", then the probability mass of $\mathcal{T}_1 \backslash \mathcal{T}_2$ is strictly positive. At the same time, $\mathcal{T}_2 \backslash \mathcal{T}_1 = \mathcal{T}_{2a} \cup \mathcal{T}_{2b} \cup \mathcal{T}_{2c}$ with

$$\begin{aligned}
\mathcal{T}_{2a} &= (\tilde{z}_1, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}] \\
\mathcal{T}_{2b} &= [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2] \times (\tilde{z}_3 + \delta_{3A}, \tilde{z}_3 + \Delta_{3A}] \\
\mathcal{T}_{2c} &= (\tilde{z}_1, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2] \times (\tilde{z}_3 + \delta_{3A}, \tilde{z}_3 + \Delta_{3A}].
\end{aligned}$$

$\mathcal{T}_{2b}$ is in $\tilde{z} + \mathcal{O}_{--+}$ and, thus, has the zero probability. If both $\mathcal{T}_{2a}$ and $\mathcal{T}_{2c}$ have the zero probability mass, then this gives us a contradiction with the supposition that both $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ and $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ give the same observable $Q_{1,2,3}(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1 - \delta_{1A}$, $z_2 = \tilde{z}_2 - \Delta_{2A}$ and $z_3 = \tilde{z}_3$. If $\mathcal{T}_{2a}$ or $\mathcal{T}_{2c}$ has a strictly positive probability mass, then this means that necessarily $\tilde{z} + \mathcal{O}_{+-+}$ is also an "intersecting orthant" and then we can use constructions from the case we have already considered to obtain a contradiction. If $\delta_{3A} \geq \Delta_{3A}$, then $\mathcal{T}_1 \backslash \mathcal{T}_2 = \mathcal{T}_{1a} \cup \mathcal{T}_{1b}$, where

$$\begin{aligned}
\mathcal{T}_{1a} &= [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times (\tilde{z}_2, \tilde{z}_2 - \Delta_{2A} + \delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}], \\
\mathcal{T}_{1b} &= [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2 - \Delta_{2A} + \delta_{2A}] \times (\tilde{z}_3 + \Delta_{3A}, \tilde{z}_3 + \delta_{3A}].
\end{aligned}$$

Rectangle $\mathcal{T}_{1a}$ is in $\tilde{z} + \mathcal{O}_{-++}$ and has a strictly positive probability mass, thus implying that the probability mass of $\mathcal{T}_1 \backslash \mathcal{T}_2$ is strictly positive. At the same time,

$$\mathcal{T}_2 \backslash \mathcal{T}_1 = (\tilde{z}_1, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}]$$

is in $\tilde{z} + \mathcal{O}_{+-+}$. If it has the zero probability mass, then this immediately gives us a contradiction with the supposition that both $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ and $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ give the same observable $Q(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1 - \delta_{1A}$, $z_2 = \tilde{z}_2 - \Delta_{2A}$ and $z_3 = \tilde{z}_3$. If $\mathcal{T}_2 \backslash \mathcal{T}_1$ has a strictly positive probability mass, then this means that necessarily $\mathcal{O}_{+-+}$ is also an "intersecting orthant" and then we can use constructions from the case we have already considered to obtain a contradiction.

(*Situation S2*) Suppose $\mathcal{O}_{-++}$ is an "empty orthant".

First, suppose that $\mathcal{O}_{+++}$ is an "intersecting orthant", which is adjacent to $\mathcal{O}_{-++}$. Then the uniqueness of the thresholds can be proven by continuing to employ the identification in the direction of the orthant

$\mathcal{O}_{+++}$. Consider[30] the following two 3-dimensional rectangles:

$$\begin{aligned}\mathcal{T}_1 &= [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times [\tilde{z}_2, \tilde{z}_2 + \delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]\\ \mathcal{T}_2 &= [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times [\tilde{z}_2, \tilde{z}_2 + \Delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}].\end{aligned}$$

If $\delta_{3A} < \Delta_{3A}$, then

$$\mathcal{T}_1 \backslash \mathcal{T}_2 = [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times (\tilde{z}_2 + \Delta_{2A}, \tilde{z}_2 + \delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]$$

is in $\tilde{z} + \mathcal{O}_{-++}$ and, thus, has zero probability, whereas $\mathcal{T}_2 \backslash \mathcal{T}_1 = \mathcal{T}_{2a} \cup \mathcal{T}_{2b}$ with

$$\begin{aligned}\mathcal{T}_{2a} &= (\tilde{z}_1, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times [\tilde{z}_2, \tilde{z}_2 + \Delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}]\\ \mathcal{T}_{2b} &= [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times [\tilde{z}_2, \tilde{z}_2 + \Delta_{2A}] \times (\tilde{z}_3 + \delta_{3A}, \tilde{z}_3 + \Delta_{3A}].\end{aligned}$$

Note that the probability mass of $\mathcal{T}_{2a}$, which is in $\tilde{z} + \mathcal{O}_{+-+}$, is strictly positive, thus implying that the probability mass of $\mathcal{T}_2 \backslash \mathcal{T}_1$ is strictly positive. This gives us a contradiction with the supposition that both $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ and $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ give the same observable $Q_{1,2,3}(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1 - \delta_{1A}$, $z_2 = \tilde{z}_2$ and $z_3 = \tilde{z}_3$. If $\delta_{3A} > \Delta_{3A}$, then $\mathcal{T}_1 \backslash \mathcal{T}_2 = \mathcal{T}_{1a} \cup \mathcal{T}_{1b}$, where

$$\begin{aligned}\mathcal{T}_{1a} &= [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times (\tilde{z}_2 + \Delta_{2A}, \tilde{z}_2 + \delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]\\ \mathcal{T}_{1b} &= [\tilde{z}_1 - \delta_{1A}, \tilde{z}_1] \times [\tilde{z}_2, \tilde{z}_2 + \delta_{2A}] \times (\tilde{z}_3 + \Delta_{3A}, \tilde{z}_3 + \delta_{3A}].\end{aligned}$$

Both $\mathcal{T}_{1a}$ and $\mathcal{T}_{1b}$ are in $\tilde{z} + \mathcal{O}_{-++}$ and, thus, have zero probability, whereas

$$\mathcal{T}_2 \backslash \mathcal{T}_1 = (\tilde{z}_1, \tilde{z}_1 - \delta_{1A} + \Delta_{1A}] \times [\tilde{z}_2, \tilde{z}_2 + \Delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}]$$

is in $\tilde{z} + \mathcal{O}_{+++}$, and since the latter is an "intersecting orthant", $\mathcal{T}_2 \backslash \mathcal{T}_1$ has a non-zero probability. Once again, this gives us a contradiction with the supposition that both $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ and $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ give the same observable $Q_{1,2,3}(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1 - \delta_{1A}$, $z_2 = \tilde{z}_2$ and $z_3 = \tilde{z}_3$.

Second, suppose $\tilde{z} + \mathcal{O}_{--+}$ is an "intersecting orthant" with $\tilde{z} + \mathcal{O}_{-++}$ continuing to be an "empty orthant". In this case we can prove identification in the direction of the orthant $\mathcal{O}_{+-+}$. In this case we suppose that there are two sets of parameters $(\underset{>0}{\Delta_{1B}}, \underset{<0}{\Delta_{2B}}, \underset{>0}{\Delta_{3B}})$ and $(\underset{>0}{\delta_{1B}}, \underset{<0}{\delta_{2B}}, \underset{>0}{\delta_{3B}})$. At least one inequality among

$$\Delta_{1B} > \delta_{1B},\ |\Delta_{2B}| > |\delta_{2B}|,\ \Delta_{3B} > \delta_{3B}$$

[30] Recall that we supposed that $\delta_{1A} < \Delta_{1A}$ and $\delta_{2A} > \Delta_{2A}$.

and at least one inequality among

$$\Delta_{1B} < \delta_{1B},\ |\Delta_{2B}| < |\delta_{2B}|,\ \Delta_{3B} < \delta_{3B}$$

must be satisfied. For concreteness, suppose $\Delta_{1B} > \delta_{1B}$ and $|\Delta_{2B}| < |\delta_{2B}|$.

Consider the following two 3-dimensional rectangles:

$$\begin{aligned}
\mathcal{T}_1 &= [\tilde{z}_1 - \delta_{1B}, \tilde{z}_1] \times [\tilde{z}_2 + |\Delta_{2B}| - |\delta_{2B}|, \tilde{z}_2 + |\Delta_{2B}|] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3B}] \\
\mathcal{T}_2 &= [\tilde{z}_1 - \delta_{1B}, \tilde{z}_1 - \delta_{1B} + \Delta_{1B}] \times [\tilde{z}_2, \tilde{z}_2 + |\Delta_{2B}|] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3B}].
\end{aligned}$$

If $\delta_{3B} < \Delta_{3B}$, then the rectangle

$$\mathcal{T}_1 \backslash \mathcal{T}_2 = [\tilde{z}_1 - \delta_{1B}, \tilde{z}_1] \times [\tilde{z}_2 + |\Delta_{2B}| - |\delta_{2B}|, \tilde{z}_2) \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3B}]$$

has a strictly positive probability given that $\tilde{z} + \mathcal{O}_{--+}$ is an "intersecting" orthant. At the same time, $\mathcal{T}_2 \backslash \mathcal{T}_1 = \mathcal{T}_{2a} \cup \mathcal{T}_{2b} \cup \mathcal{T}_{2c}$ with

$$\begin{aligned}
\mathcal{T}_{2a} &= (\tilde{z}_1, \tilde{z}_1 - \delta_{1B} + \Delta_{1B}] \times [\tilde{z}_2, \tilde{z}_2 + |\Delta_{2B}|] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3B}] \\
\mathcal{T}_{2b} &= [\tilde{z}_1 - \delta_{1B}, \tilde{z}_1] \times [\tilde{z}_2, \tilde{z}_2 + |\Delta_{2B}|] \times (\tilde{z}_3 + \delta_{3B}, \tilde{z}_3 + \Delta_{3B}] \\
\mathcal{T}_{2c} &= (\tilde{z}_1, \tilde{z}_1 - \delta_{1B} + \Delta_{1B}] \times [\tilde{z}_2, \tilde{z}_2 + |\Delta_{2B}|] \times (\tilde{z}_3 + \delta_{3B}, \tilde{z}_3 + \Delta_{3B}]
\end{aligned}$$

$\mathcal{T}_{2b}$ is in $\tilde{z} + \mathcal{O}_{-++}$ and, thus, has probability zero. If $\mathcal{T}_{2a}$ and $\mathcal{T}_{2c}$ have probability zero as well, then this immediately gives us a contradiction with the supposition that both $(\Delta_{1B}, \Delta_{2B}, \Delta_{3B})$ and $(\delta_{1B}, \delta_{2B}, \delta_{3B})$ give the same observable $Q_{1,\bar{2},3}(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1 - \delta_{1B}$, $z_2 = \tilde{z}_2 + |\Delta_{2B}|$ and $z_3 = \tilde{z}_3$. If either $\mathcal{T}_{2a}$ and $\mathcal{T}_{2c}$ has a strictly positive probability, then by convexity of $\mathcal{E}_{123}$ this would imply that $\tilde{z} + \mathcal{O}_{+++}$ is necessarily an "intersecting orthant"' and then we can use constructions from the case we have already considered to obtain a contradiction. If $\delta_{3B} \geq \Delta_{3B}$, then $\mathcal{T}_1 \backslash \mathcal{T}_2 = \mathcal{T}_{1a} \cup \mathcal{T}_{1b}$, where

$$\begin{aligned}
\mathcal{T}_{1a} &= [\tilde{z}_1 - \delta_{1B}, \tilde{z}_1] \times [\tilde{z}_2 + |\Delta_{2B}| - |\delta_{2B}|, \tilde{z}_2) \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3B}] \\
\mathcal{T}_{1b} &= [\tilde{z}_1 - \delta_{1B}, \tilde{z}_1] \times [\tilde{z}_2 + |\Delta_{2B}| - |\delta_{2B}|, \tilde{z}_2 + |\Delta_{2B}|] \times (\tilde{z}_3 + \Delta_{3B}, \tilde{z}_3 + \delta_{3B}]
\end{aligned}$$

$\mathcal{T}_{1a}$ has a strictly positive probability because $\tilde{z} + \mathcal{O}_{--+}$ is an "intersecting" orthant, thus giving an overall strictly positive probability of the whole $\mathcal{T}_1 \backslash \mathcal{T}_2$. At the same time,

$$\mathcal{T}_2 \backslash \mathcal{T}_1 = (\tilde{z}_1, \tilde{z}_1 - \delta_{1B} + \Delta_{1B}] \times [\tilde{z}_2, \tilde{z}_2 + |\Delta_{2B}|] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3B}].$$

If $\mathcal{T}_2 \backslash \mathcal{T}_1$ has probability zero, then this immediately gives us a contradiction with the supposition that both $(\Delta_{1B}, \Delta_{2B}, \Delta_{3B})$ and $(\delta_{1B}, \delta_{2B}, \delta_{3B})$ give the same observable $Q_{1,\overline{2},3}(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1 - \delta_{1B}$, $z_2 = \tilde{z}_2 + |\Delta_{2B}|$ and $z_3 = \tilde{z}_3$. If either $\mathcal{T}_2 \backslash \mathcal{T}_1$ has a strictly positive probability, then by convexity of $\mathcal{E}_{123}$ this would imply that $\tilde{z} + \mathcal{O}_{+++}$ is necessarily an "intersecting orthant" and then we can use constructions from the case we have already considered to obtain a contradiction.

(*Situation S3*) Suppose $\mathcal{O}_{+-+}$ is an "empty orthant".

First, suppose that $\tilde{z} + \mathcal{O}_{+++}$ is an "intersecting orthant", which is adjacent to $\mathcal{O}_{+-+}$. Then the uniqueness of the thresholds can be proven by continuing to employ the identification in the direction of the orthant $\mathcal{O}_{+++}$. Consider[31] the following two 3-dimensional rectangles:

$$\mathcal{T}_1 = [\tilde{z}_1, \tilde{z}_1 + \delta_{1A}] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2 - \Delta_{2A} + \delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]$$
$$\mathcal{T}_2 = [\tilde{z}_1, \tilde{z}_1 + \Delta_{1A}] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}].$$

If $\delta_{3A} < \Delta_{3A}$, then

$$\mathcal{T}_1 \backslash \mathcal{T}_2 = [\tilde{z}_1, \tilde{z}_1 + \delta_{1A}] \times (\tilde{z}_2, \tilde{z}_2 - \Delta_{2A} + \delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]$$

has a strictly positive probability mass as $\tilde{z} + \mathcal{O}_{+++}$ is an "intersecting orthant". Also, $\mathcal{T}_2 \backslash \mathcal{T}_1 = \mathcal{T}_{2a} \cup \mathcal{T}_{2b}$ with

$$\mathcal{T}_{2a} = (\tilde{z}_1 + \delta_{1A}, \tilde{z}_1 + \Delta_{1A}] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}]$$
$$\mathcal{T}_{2b} = [\tilde{z}_1, \tilde{z}_1 \Delta_{1A}] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2] \times (\tilde{z}_3 + \delta_{3A}, \tilde{z}_3 + \Delta_{3A}].$$

Both $\mathcal{T}_{2a}$ and $\mathcal{T}_{2b}$ are in $\tilde{z} + \mathcal{O}_{+-+}$, and, thus, have zero probability mass. This gives us a contradiction with the supposition that both $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ and $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ give the same observable $Q(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1$, $z_2 = \tilde{z}_2 - \Delta_{2A}$ and $z_3 = \tilde{z}_3$.

If $\delta_{3A} > \Delta_{3A}$, then $\mathcal{T}_1 \backslash \mathcal{T}_2 = \mathcal{T}_{1a} \cup \mathcal{T}_{1b}$, where

$$\mathcal{T}_{1a} = [\tilde{z}_1, \tilde{z}_1 + \delta_{1A}] \times (\tilde{z}_2, \tilde{z}_2 - \Delta_{2A} + \delta_{2A}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3A}]$$
$$\mathcal{T}_{1b} = [\tilde{z}_1, \tilde{z}_1 + \delta_{1A}] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2 - \Delta_{2A} + \delta_{2A}] \times (\tilde{z}_3 + \Delta_{3A}, \tilde{z}_3 + \delta_{3A}].$$

$\mathcal{T}_1 \backslash \mathcal{T}_2$ has a strictly positive probability mass as $\mathcal{T}_{1a}$ has a strictly positive probability mass because of

[31] Recall that we supposed that $\delta_{1A} < \Delta_{1A}$ and $\delta_{2A} > \Delta_{2A}$.

$\tilde{z} + \mathcal{O}_{+++}$ being an "intersecting orthant".. At the same time,

$$\mathcal{T}_2 \backslash \mathcal{T}_1 = (\tilde{z}_1 + \delta_{1A}, \tilde{z}_1 + \Delta_{1A}] \times [\tilde{z}_2 - \Delta_{2A}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3A}]$$

is in $\tilde{z} + \mathcal{O}_{+-+}$ and has the probability mass of zero since $\tilde{z} + \mathcal{O}_{+-+}$ is an "empty orthant". Once again, this gives us a contradiction with the supposition that both $(\Delta_{1A}, \Delta_{2A}, \Delta_{3A})$ and $(\delta_{1A}, \delta_{2A}, \delta_{3A})$ give the same observable $Q_{1,2,3}(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1$, $z_2 = \tilde{z}_2 - \Delta_{2A}$ and $z_3 = \tilde{z}_3$.

Second, suppose $\tilde{z} + \mathcal{O}_{--+}$ is an "intersecting orthant" with $\tilde{z} + \mathcal{O}_{+-+}$ continuing to be an "empty orthant". In this case we can prove identification in the direction of the orthant $\mathcal{O}_{-++}$. In this case we suppose that there are two sets of parameters $(\underset{<0}{\Delta_{1C}}, \underset{>0}{\Delta_{2Cc}}, \underset{>0}{\Delta_{3C}})$ and $(\underset{<0}{\delta_{1C}}, \underset{>0}{\delta_{2Cc}}, \underset{>0}{\delta_{3C}})$. At least one inequality among

$$|\Delta_{1C}| > |\delta_{1C}|,\ \Delta_{2C} > \delta_{2C},\ \Delta_{3C} > \delta_{3C}$$

and at least one inequality among

$$|\Delta_{1C}| < |\delta_{1C}|,\ \Delta_{2C} < \delta_{2C},\ \Delta_{3C} < \delta_{3C}$$

must be satisfied. For concreteness, suppose $|\Delta_{1C}| > |\delta_{1C}|$ and $\Delta_{2C} < \delta_{2C}$.

Consider the following two 3-dimensional rectangles:

$$\mathcal{T}_1 = [\tilde{z}_1, \tilde{z}_1 + |\delta_{1C}|] \times [\tilde{z}_2 - \Delta_{2C}, \tilde{z}_2 - \Delta_{2C} + \delta_{2C}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3C}]$$
$$\mathcal{T}_2 = [\tilde{z}_1 + |\delta_{1C}| - |\Delta_{1C}|, \tilde{z}_1 + |\delta_{1C}|] \times [\tilde{z}_2 - \Delta_{2C}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3C}].$$

If $\delta_{3C} < \Delta_{3C}$, then $\mathcal{T}_2 \backslash \mathcal{T}_1 = \mathcal{T}_{2a} \cup \mathcal{T}_{2b}$ with

$$\mathcal{T}_{2a} = [\tilde{z}_1 + |\delta_{1C}| - |\Delta_{1C}|, \tilde{z}_1) \times [\tilde{z}_2 - \Delta_{2C}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3C}],$$
$$\mathcal{T}_{2b} = [\tilde{z}_1 + |\delta_{1C}| - |\Delta_{1C}|, \tilde{z}_1 + |\delta_{1C}|] \times [\tilde{z}_2 - \Delta_{2C}, \tilde{z}_2] \times (\tilde{z}_3 + \delta_{3C}, \tilde{z}_3 + \Delta_{3C}].$$

$\mathcal{T}_{2a}$ has a strictly positive probability mass since $\tilde{z} + \mathcal{O}_{--+}$ is an "intersecting orthant", thus implying a strictly positive probability mass of the whole $\mathcal{T}_2 \backslash \mathcal{T}_1$. At the same time,

$$\mathcal{T}_1 \backslash \mathcal{T}_2 = [\tilde{z}_1, \tilde{z}_1 + |\delta_{1C}|] \times (\tilde{z}_2, \tilde{z}_2 - \Delta_{2C} + \delta_{2C}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3C}].$$

If $\mathcal{T}_1 \backslash \mathcal{T}_2$ has the probability mass of zero, then this immediately gives us a contradiction with the supposition that both $(\Delta_{1C}, \Delta_{2C}, \Delta_{3C})$ and $(\delta_{1C}, \delta_{2C}, \delta_{3C})$ give the same observable $Q_{\bar{1},2,3}(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1 + |\delta_{1C}|$, $z_2 = \tilde{z}_2 + |\Delta_{2C}|$ and $z_3 = \tilde{z}_3$. If $\mathcal{T}_1 \backslash \mathcal{T}_2$ has a strictly positive probability, then by convexity of $\mathcal{E}_{123}$ this would imply that $\tilde{z} + \mathcal{O}_{+++}$ is necessarily an "intersecting orthant"' and then we can use constructions from the case we have already considered to obtain a contradiction.

If $\delta_{3C} \geq \Delta_{3C}$, then the rectangle

$$\mathcal{T}_2 \backslash \mathcal{T}_1 = [\tilde{z}_1 + |\delta_{1C}| - |\Delta_{1C}|, \tilde{z}_1) \times [\tilde{z}_2 - \Delta_{2C}, \tilde{z}_2] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3C}]$$

has a strictly positive probability mass since $\tilde{z} + \mathcal{O}_{--+}$ is an "intersecting orthant". At the same time, $\mathcal{T}_1 \backslash \mathcal{T}_2 = \mathcal{T}_{1a} \cup \mathcal{T}_{1b} \cup \mathcal{T}_{1c}$ with

$$\begin{aligned}
\mathcal{T}_{1a} &= [\tilde{z}_1, \tilde{z}_1 + |\delta_{1C}|] \times (\tilde{z}_2, \tilde{z}_2 - \Delta_{2C} + \delta_{2C}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3C}], \\
\mathcal{T}_{1b} &= [\tilde{z}_1, \tilde{z}_1 + |\delta_{1C}|] \times [\tilde{z}_2 - \Delta_{2C}, \tilde{z}_2] \times (\tilde{z}_3 + \Delta_{3C}, \tilde{z}_3 + \delta_{3C}], \\
\mathcal{T}_{1c} &= [\tilde{z}_1, \tilde{z}_1 + |\delta_{1C}|] \times (\tilde{z}_2, \tilde{z}_2 - \Delta_{2C} + \delta_{2C}] \times (\tilde{z}_3 + \Delta_{3C}, \tilde{z}_3 + \delta_{3C}].
\end{aligned}$$

$\mathcal{T}_{1b}$ is in $\tilde{z} + \mathcal{O}_{+-+}$ and, thus, has probability zero. If $\mathcal{T}_{1a}$ and $\mathcal{T}_{1c}$ have probability zero as well, then this immediately gives us a contradiction with the supposition that both $(\Delta_{1C}, \Delta_{2C}, \Delta_{3C})$ and $(\delta_{1C}, \delta_{2C}, \delta_{3C})$ give the same observable $Q_{1,\overline{2},3}(z_1, z_2, z_3)$ if we take $z_1 = \tilde{z}_1 + |\delta_{1C}|$, $z_2 = \tilde{z}_2 + |\Delta_{2C}|$ and $z_3 = \tilde{z}_3$. If either $\mathcal{T}_{1a}$ and $\mathcal{T}_{1c}$ has a strictly positive probability, then by convexity of $\mathcal{E}_{123}$ this would imply that $\tilde{z} + \mathcal{O}_{+++}$ is necessarily an "intersecting orthant"' and then we can use constructions from the case we have already considered to obtain a contradiction.

(*Situation S4*) The final case is when $\mathcal{O}_{+++}$ is an "empty orthant".

If we, first, suppose that $\tilde{z} + \mathcal{O}_{+-+}$ is an "intersecting orthant", which is adjacent to $\mathcal{O}_{+++}$, then we can prove identification in the direction of the orthant $\mathcal{O}_{+-+}$ similar to how it was done in situation S2. In this case we suppose that there are two sets of parameters $(\underset{>0}{\Delta_{1B}}, \underset{<0}{\Delta_{2B}}, \underset{>0}{\Delta_{3B}})$ and $(\underset{>0}{\delta_{1B}}, \underset{<0}{\delta_{2B}}, \underset{>0}{\delta_{3B}})$. For concreteness, we can suppose $\Delta_{1B} > \delta_{1B}$ and $|\Delta_{2B}| < |\delta_{2B}|$. Then we can obtain contradictions by considering the following two 3-dimensional rectangles:

$$\begin{aligned}
\mathcal{T}_1 &= [\tilde{z}_1, \tilde{z}_1 + \delta_{1B}] \times [\tilde{z}_2 + |\Delta_{2B}| - |\delta_{2B}|, \tilde{z}_2 + |\Delta_{2B}|] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3B}] \\
\mathcal{T}_2 &= [\tilde{z}_1, \tilde{z}_1 + \Delta_{1B}] \times [\tilde{z}_2, \tilde{z}_2 + |\Delta_{2B}|] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3B}].
\end{aligned}$$

Namely, $\mathcal{T}_1 \backslash \mathcal{T}_2$ will have a strictly positive probability mass whereas $\mathcal{T}_2 \backslash \mathcal{T}_1$ will have the probability mass of zero.

If we, second, suppose that $\tilde{z} + \mathcal{O}_{-++}$ is an "intersecting orthant", which is adjacent to $\mathcal{O}_{+++}$, then we can prove identification in the direction of the orthant $\mathcal{O}_{-++}$ similar to how it was done in situation S3. In this case we suppose that there are two sets of parameters $(\underset{<0}{\Delta_{1C}}, \underset{>0}{\Delta_{2Cc}}, \underset{>0}{\Delta_{3C}})$ and $(\underset{<0}{\delta_{1C}}, \underset{>0}{\delta_{2Cc}}, \underset{>0}{\delta_{3C}})$. For concreteness, we can suppose $|\Delta_{1C}| > |\delta_{1C}|$ and $\Delta_{2C} < \delta_{2C}$. Then we can obtain contradictions by considering the following two 3-dimensional rectangles:

$$\mathcal{T}_1 = [\tilde{z}_1, \tilde{z}_1 + |\delta_{1C}|] \times [\tilde{z}_2, \tilde{z}_2 + \delta_{2C}] \times [\tilde{z}_3, \tilde{z}_3 + \delta_{3C}]$$
$$\mathcal{T}_2 = [\tilde{z}_1 + |\delta_{1C}| - |\Delta_{1C}|, \tilde{z}_1 + |\delta_{1C}|] \times [\tilde{z}_2, \tilde{z}_2 + \Delta_{2C}] \times [\tilde{z}_3, \tilde{z}_3 + \Delta_{3C}].$$

Namely, $\mathcal{T}_1 \backslash \mathcal{T}_2$ will have the probability mass of zero whereas $\mathcal{T}_2 \backslash \mathcal{T}_1$ will have a strictly positive probability mass.

All the contradictions obtained in situations S1-S4 are derived with a strictly positive probability since the difference in the described regions will continue to hold for points at the intersection of the interior of $\mathcal{E}_1 23$ and a small neighborhood of the boundary point $\tilde{z}$.

All the subsequent steps of showing uniqueness of thresholds when the indices in four dimensions change and so n can be shown analogously to Step 6. $\square$

C Identification in parametric models

C.1 Lattice ordered probit

Here we show the identification of an ordered probit model with a lattice structure. Identification is split into two parts. The first part in Theorem 8 gives identification of the index parameters β_d, $d = 1, \ldots, D$. The second part gives sufficient conditions on the identification of the correlation coefficient ρ_{d_1,d_2} given that both β_{d_1} and β_{d_2} are identified.

Theorem 8 (identification of the index parameter) *Suppose Assumption 5 holds. If for dimension d, there are $k_d + 1$ points $\{x_d^{(i)}\}_{i=1}^{k_d+1}$ in $\mathcal{X}_d$ such that the matrix*

$$\begin{pmatrix} 1 & x_d^{(1)} \\ 1 & x_d^{(2)} \\ \vdots & \vdots \\ 1 & x_d^{(k_d+1)} \end{pmatrix}$$

has rank $k_d + 1$, then β_d and $\{\alpha_j^{(d)}\}$ is identified.

The main condition in Theorem 8 is simply the rank condition or the condition on a sufficient variation in covariates in dimension d. The identification of correlation coefficients can be conducted in a pairwise fashion due to the lattice structure of the model. Theorem 9 gives various sufficient conditions for identifying the correlation coefficients.

Proof of Theorem 8. *Let $\Phi(\cdot)$ denote the c.d.f. of the standard normal distribution. Then for any $d = 1, \ldots, D$ and for any $j = 1, \ldots, M_d - 1$, we can use the lattice structure of thresholds to obtain*

$$P\left(Y^{(d)} \leq y_j^{(d)} | x_1, \ldots, x_D\right) = P\left(Y^{(d)} \leq y_j^{(d)} | x_d\right) = \Phi\left(\alpha_j^{(d)} - x_d\beta_d\right),$$

and hence,

$$\Phi^{-1}\left(P\left(Y^{(d)} \leq y_j^{(d)} | x_d\right)\right) = \alpha_j^{(d)} - x_d\beta_d,$$

where the left-hand side is known from the distribution of observables.

The condition in the theorem allows us to construct a system of $k_d + 1$ linear equations with $k_d + 1$ unknowns in $(\alpha_j^{(d)}, \beta_d)$ whose system of coefficients has full rank, thus implying the identification of $(\alpha_j^{(d)}, \beta_d)$. □

Theorem 9 (Identification of correlation coefficients) *Suppose Assumption 5 holds and conditions of Theorem 8 hold for dimensions d_1 and d_2, $d_1 \neq d_2$. Then the correlation coefficient ρ_{d_1,d_2} is identified if at least one of the following conditions hold:*

(a) there is a point $x^*_{d_1} \in \mathcal{X}_{d_1}$ such that $\alpha_j^{(d_1)} - x^*_{d_1}\beta_{d_1} = 0$ for some $j = 1, \ldots, M_{d_1}$;

(b) At least three different rectangular regions $\mathcal{I}_{j_{d_1}}^{(d_1)} \times \mathcal{I}_{j_{d_2}}^{(d_2)}$ (see definition in (4)) contain points $(x_{d_1}\beta_{d_1}, x_{d_2}\beta_{d_2})$ from some $(x_{d_1}, x_{d_2}) \in \mathcal{X}_{d_1 d_2}$.

(c) There are variables in x_{d_1} – without a loss of generality suppose they form a subvector $x_{d_1,1:L_{d_1}}$, $L_{d_1} \geq 1$, – such that at least of the parameters in $\beta_{d_1,1:L_{d_1}}$ is not zero and and $x_{d_1,1:L_{d_1}}$ is excluded from x_{d2} – that is,

$$x_{d_1,\ell}\,|\,x_{d_2} \text{ has a non-degenerate distribution}, \quad l = 1, \ldots, L_{d_1}.$$

There are two different points in $\mathcal{X}_{d_1 d_2}$ that differ only in the value of covariates in the subvector $x_{d_1,1:L_{d_1}}$ – denote them as $(x^{(h)}_{d_1,1:L_{d_1}}, x_{d_1,L_{d_1}+1:k_{d_1}}, x_{d_2})$, $h = 1, 2$, such that for some index $j_{d_1} \leq M_{d_1} - 1$

$$P\left(Y^{(d_1)} \leq y_{j_1}^{(d_1)}, Y^{(d_2)} \leq y_{j_2}^{(d_2)}\,|\,x^{(1)}_{d_1,1:L_{d_1}}, x_{d_1,L_{d_1}+1:k_{d_1}}, x_{d_2}\right) \neq$$
$$P\left(Y^{(d_1)} \leq y_{j_1}^{(d_1)}, Y^{(d_2)} \leq y_{j_2}^{(d_2)}\,|\,x^{(2)}_{d_1,1:L_{d_1}}, x_{d_1,L_{d_1}+1:k_{d_1}}, x_{d_2}\right).$$

Proof of Theorem 9. *(a) Take j_1 such that $\alpha^{(d_1)}_{j_1} - x^*_{d_1}\beta_{d_1} = 0$. Find the whole vector x^* that has $x^*_{d_1}$ as a vector of covariates in the d_1-th process, and extract $x^*_{d_2}$ from x^*. If $\alpha^{(d_2)}_{j_2} - x^*_{d_2}\beta_{d_2} \leq 0$, consider the known probability*

$$P\left(Y^{(d_1)} \leq y^{(d_1)}_{j_1},\ Y^{(d_2)} \leq y^{(d_2)}_{j_2} \,|\, x^*_{d_1}, x^*_{d_2}\right) = \int_{-\infty}^{\alpha^{(d_2)}_{j_2} - x^*_{d_2}\beta_{d_2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(-\frac{\rho_{d_1,d_2}}{\sqrt{1-\rho^2_{d_1,d_2}}}\eta\right) d\eta.$$

Because $\alpha^{(d_2)}_{j_2} - x^*_{d_2}\beta_{d_2} \leq 0$, the right-hand side is strictly increasing in $\frac{\rho_{d_1,d_2}}{\sqrt{1-\rho^2_{d_1,d_2}}}$ and everything else on the right-hand side is known. Therefore, $\frac{\rho_{d_1,d_2}}{\sqrt{1-\rho^2_{d_1,d_2}}}$ is identified. Since $\frac{\rho_{d_1,d_2}}{\sqrt{1-\rho^2_{d_1,d_2}}}$ in its turn is a strictly increasing function of $\rho_{d_1,d_2} \in (-1,1)$, this guarantees that identification of ρ_{d_1,d_2}. If $\alpha^{(d_2)}_{j_2} - x^*_{d_2}\beta_{d_2} < 0$, then instead we would consider the probability $P\left(Y^{(d_1)} \leq y^{(d_1)}_{j_1},\ Y^{(d_2)} > y^{(d_2)}_{j_2} \,|\, x^*_{d_1}, x^*_{d_2}\right)$ and conduct an analogous identification strategy.

(b) The condition implies that there are indices j_1^0 and j_2^0 such as at least three of the following four systems (48)-(51) of inequalities have a solution $(x_{d_1}, x_{d_2}) \in \mathcal{X}_{d_1,d_2}$:

$$\alpha^{(d_1)}_{j_1^0} - x_{d_1}\beta_{d_1} \geq 0, \quad \alpha^{(d_2)}_{j_2^0} - x_{d_2}\beta_{d_2} \geq 0, \tag{48}$$

$$\alpha^{(d_1)}_{j_1^0} - x_{d_1}\beta_{d_1} \geq 0, \quad \alpha^{(d_2)}_{j_2^0} - x_{d_2}\beta_{d_2} < 0, \tag{49}$$

$$\alpha^{(d_1)}_{j_1^0} - x_{d_1}\beta_{d_1} < 0, \quad \alpha^{(d_2)}_{j_2^0} - x_{d_2}\beta_{d_2} \geq 0, \tag{50}$$

$$\alpha^{(d_1)}_{j_1^0} - x_{d_1}\beta_{d_1} < 0, \quad \alpha^{(d_2)}_{j_2^0} - x_{d_2}\beta_{d_2} < 0. \tag{51}$$

Note that which exactly three systems among (48)-(51) have solutions only determines which probabilities we consider below. For the sake of expositional simplicity, introduce generic notations $c_{d_1} = \alpha^{(d_1)}_{j_1^0} - x_{d_1}\beta_{d_1}$ and $c_{d_2} = \alpha^{(d_2)}_{j_2^0} - x_{d_2}\beta_{d_2}$.

Here is the outline of the identification strategy. Among three systems with solutions, we can find two systems such that in one system both c_{d_1} and c_{d_2} have the same sign and in the other system one of c_{d_1} and c_{d_2} preserves the same sign as in the first system. Suppose, without a loss of generality it is c_{d_2} that has the same sign in both systems. If c_{d_2} in both systems is positive, we consider conditional probabilities of $\left\{Y^{(d_1)} \leq y^{(d_1)}_{j_1^0}, Y^{(d_2)} > y^{(d_2)}_{j_2^0}\right\}$ for points that satisfy those two systems. If c_{d_2} in both systems is negative, we consider conditional probabilities of $\left\{Y^{(d_1)} \leq y^{(d_1)}_{j_1^0}, Y^{(d_2)} \leq y^{(d_2)}_{j_2^0}\right\}$ for points that satisfy those two systems. in either case, we will be able to conclude that among non-negative ρ_{d_1,d_2} at most one values can generate observables, and similarly, among non-positive ρ_{d_1,d_2} at most one values can generate observables. Thus, it is possible to have at most two values of ρ_{d_1,d_2} of different signs.

Now, in the third system it is guaranteed that c_{d_2} will have a sign opposite to the sign in the first two *systems. This can be used to establish a strict inequality between the absolute values of two possibly compatible different ρ_{d_1,d_2}. Then, going back to one of the first two systems where the sign of c_{d_1} is the same as in the third system, we will be able to establish a strict inequality between the absolute values of two possibly compatible different ρ_{d_1,d_2} which will contradict the inequality obtained from the previous step. This contradiction will allow us to conclude that there can be only one ρ_{d_1,d_2}.*

To make this discussion more specific, consider e.g. the case when systems (48), (50) and (51) have solutions. Following the outline of the identification strategy above, we can consider (48) and (50) as the first two systems. In both these systems c_{d_2} is non-negative. The immediate implication is that c_{d_1} has different signs in these two systems. Let us show how to utilize this.

Take a point $\left(x^{(1)}_{d_1}, x^{(1)}_{d_2}\right)$ that satisfies (48). Then on the right-hand side of

$$P\left(Y^{(d_1)} \leq y^{(d_1)}_{j_1^0}, Y^{(d_2)} > y^{(d_2)}_{j_2^0} \,|\, x^{(1)}_{d_1}, x^{(1)}_{d_2}\right) = \int_{\alpha^{(d_2)}_{j_2^0} - x^{(1)}_{d_2}\beta_{d_2}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(\frac{\alpha^{(d_1)}_{j_1^0} - x^{(1)}_{d_1}\beta_{d_1} - \rho_{d_1,d_2}\eta}{\sqrt{1-\rho^2_{d_1,d_2}}}\right) d\eta,$$

the only unknown component is ρ_{d_1,d_2} (see Theorem 8) and $-\frac{\rho_{d_1,d_2}\eta}{\sqrt{1-\rho^2_{d_1,d_2}}}$ is strictly decreasing in ρ_{d_1,d_2}. Since $\alpha^{(d_1)}_{j_1^0} - x^{(1)}_{d_1}\beta_{d_1} \geq 0$, then $\frac{\alpha^{(d_1)}_{j_1^0} - x^{(1)}_{d_1}\beta_{d_1}}{\sqrt{1-\rho^2_{d_1,d_2}}}$ as a function of ρ_{d_1,d_2} is decreasing on the interval $(-1, 0]$. Hence, the whole right-hand of this probability expression is strictly decreasing in ρ_{d_1,d_2} on the interval $(-1, 0]$. Thus, among non-positive ρ_{d_1,d_2}, there can be at most one value that can generate observable left-hand side.

Now take a point $\left(x^{(2)}_{d_1}, x^{(2)}_{d_2}\right)$ that satisfies (50). Since $\alpha^{(d_1)}_{j_1^0} - x^{(2)}_{d_1}\beta_{d_1} < 0$, then analogously to above it can be concluded that the right-hand side of

$$P\left(Y^{(d_1)} \leq y^{(d_1)}_{j_1^0}, Y^{(d_2)} > y^{(d_2)}_{j_2^0} \,|\, x^{(2)}_{d_1}, x^{(2)}_{d_2}\right) = \int_{\alpha^{(d_2)}_{j_2^0} - x^{(2)}_{d_2}\beta_{d_2}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(\frac{\alpha^{(d_1)}_{j_1^0} - x^{(2)}_{d_1}\beta_{d_1} - \rho_{d_1,d_2}\eta}{\sqrt{1-\rho^2_{d_1,d_2}}}\right) d\eta$$

is strictly decreasing in ρ_{d_1,d_2} on the interval $[0, 1)$. Hence, among non-negative ρ_{d_1,d_2}, there can be at most one value that can generate observables. By just considering these two points, we can conclude that there can be at most two values (one non-negative and one non-positive) in the identified set. Let us denote these two candidate values as $\rho^*_{d_1,d_2} \leq 0$ and $\tilde{\rho}_{d_1,d_2} > 0$.

We want to show that only of these is consistent with the data. Suppose that contrary to this both $\rho^*_{d_1,d_2}$ and $\tilde{\rho}_{d_1,d_2}$ can generate observables. Following the identification strategy outlined as above, we now take *a point* $(x^{(3)}_{d_1}, x^{(3)}_{d_2})$ *that satisfies (51) and consider*[32]

$$P\left(Y^{(d_1)} \leq y^{(d_1)}_{j^0_1}, Y^{(d_2)} \leq y^{(d_2)}_{j^0_2} \mid x^{(3)}_{d_1}, x^{(3)}_{d_2}\right) = \int_{-\infty}^{\alpha^{(d_2)}_{j^0_2} - x^{(3)}_{d_2}\beta_{d_2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(\frac{\alpha^{(d_1)}_{j^0_1} - x^{(3)}_{d_1}\beta_{d_1} - \rho_{d_1,d_2}\eta}{\sqrt{1-\rho^2_{d_1,d_2}}}\right) d\eta.$$

Note that since $\alpha^{(d_2)}_{j^0_2} - x^{(3)}_{d_2}\beta_{d_2} < 0$*, the equation*

$$P\left(Y^{(d_1)} \leq y^{(d_1)}_{j^0_1}, Y^{(d_2)} \leq y^{(d_2)}_{j^0_2} \mid x^{(3)}_{d_1}, x^{(3)}_{d_2}\right) = \int_{-\infty}^{\alpha^{(d_2)}_{j^0_2} - x^{(3)}_{d_2}\beta_{d_2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(b - a\eta\right) d\eta$$

considered for all observationally equivalent (a, b)*, delivers a strictly decreasing in a function* $b(a)$ *that generates the same* $P\left(Y^{(d_1)} \leq y^{(d_1)}_{j^0_1}, Y^{(d_2)} \leq y^{(d_2)}_{j^0_2} \mid x^{(3)}_{d_1}, x^{(3)}_{d_2}\right)$*. It is easy to see that for both* $a^* = \frac{\rho^*_{d_1,d_2}}{\sqrt{1-\rho^*_{d_1,d_2}{}^2}} \leq 0$*,* $b^* = \frac{\alpha^{(d_1)}_{j^0_1} - x^{(3)}_{d_1}\beta_{d_1}}{\sqrt{1-\rho^*_{d_1,d_2}{}^2}} < 0$ *and* $\tilde{a} = \frac{\tilde{\rho}_{d_1,d_2}}{\sqrt{1-\tilde{\rho}^2_{d_1,d_2}}} > 0$*,* $\tilde{b} = \frac{\alpha^{(d_1)}_{j^0_1} - x^{(3)}_{d_1}\beta_{d_1}}{\sqrt{1-\tilde{\rho}^2_{d_1,d_2}}} < 0$ *to be compatible with the fact that they belong long to the curve* $(a, b(a))$ *with the strictly decreasing* $b(\cdot)$*, it has to be satisfied that* $|\tilde{\rho}_{d_1,d_2}| > |\rho^*_{d_1,d_2}|$*.*

Now go back to $\left(x^{(2)}_{d_1}, x^{(2)}_{d_2}\right)$ *that satisfies (50) but this time consider the probability*

$$P\left(Y^{(d_1)} \leq y^{(d_1)}_{j^0_1}, Y^{(d_2)} \leq y^{(d_2)}_{j^0_2} \mid x^{(2)}_{d_1}, x^{(2)}_{d_2}\right) = \int_{-\infty}^{\alpha^{(d_1)}_{j^0_1} - x^{(2)}_{d_1}\beta_{d_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(\frac{\alpha^{(d_2)}_{j^0_2} - x^{(2)}_{d_2}\beta_{d_2} - \rho\eta}{\sqrt{1-\rho^2}}\right) d\eta$$

Note that since $\alpha^{(d_1)}_{j^0_1} - x^{(2)}_{d_1}\beta_{d_1} < 0$*, the equation*

$$P\left(Y^{(d_1)} \leq y^{(d_1)}_{j^0_1}, Y^{(d_2)} \leq y^{(d_2)}_{j^0_2} \mid x^{(2)}_{d_1}, x^{(2)}_{d_2}\right) = \int_{-\infty}^{\alpha^{(d_1)}_{j^0_1} - x^{(2)}_{d_1}\beta_{d_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(b - a\eta\right) d\eta$$

considered for all observationally equivalent (a, b)*, delivers a strictly decreasing in a function* $b(a)$ *that generates the same* $P\left(Y^{(d_1)} \leq y^{(d_1)}_{j^0_1}, Y^{(d_2)} \leq y^{(d_2)}_{j^0_2} \mid x^{(2)}_{d_1}, x^{(2)}_{d_2}\right)$*. It is easy to see that for both* $a^* = \frac{\rho^*_{d_1,d_2}}{\sqrt{1-\rho^*_{d_1,d_2}{}^2}} \leq 0$*,* $b^* = \frac{\alpha^{(d_1)}_{j^0_1} - x^{(2)}_{d_1}\beta_{d_1}}{\sqrt{1-\rho^*_{d_1,d_2}{}^2}} > 0$ *and* $\tilde{a} = \frac{\tilde{\rho}_{d_1,d_2}}{\sqrt{1-\tilde{\rho}^2_{d_1,d_2}}} > 0$*,* $\tilde{b} = \frac{\alpha^{(d_1)}_{j^0_1} - x^{(2)}_{d_1}\beta_{d_1}}{\sqrt{1-\tilde{\rho}^2_{d_1,d_2}}} > 0$ *to be compatible with the fact that they belong long to the curve* $(a, b(a))$ *with the strictly decreasing* $b(\cdot)$*, it has to be satisfied that* $|\tilde{\rho}_{d_1,d_2}| < |\rho^*_{d_1,d_2}|$*. This is a contradiction with the previous conclusion. Therefore, only one of* $\rho^*_{d_1,d_2}$ *and* $\tilde{\rho}_{d_1,d_2}$ *can generate observables.*

(c) Denote $x^{(1)}_{d_1} = (x^{(1)}_{d_1,1:L_1}, x_{d_1,L_{d_1}+1:k_{d_1}})$ *and* $x^{(2)}_{d_1} = (x^{(2)}_{d_1,1:L_1}, x_{d_1,L_{d_1}+1:k_{d_1}})$*.*

We first consider the case when $\alpha^{d_1}_{j_{d_1}} - x^{(1)}_{d_1}\beta_{d_1}$ *and* $\alpha^{d_1}_{j_{d_1}} - x^{(2)}_{d_1}\beta_{d_1}$ *take different signs – e.g. suppose that*

[32] Note that now we consider $Y^{(d_2)} \leq y^{(d_2)}_{j^0_2}$ since now c_{d_2} has the negative sign.

$\alpha^{d_1}_{j_{d_1}} - x^{(1)}_{d_1}\beta_{d_1} \geq 0$ *and* $\alpha^{d_1}_{j_{d_1}} - x^{(2)}_{d_1}\beta_{d_1} \leq 0.$

For index j_{d_1} *in this condition and for any index* j_{d_2}*,* $j_{d_2} \leq M_{d_2} - 1$*, consider the probability*

$$P\left(Y^{(d_1)} \leq y^{(d_1)}_{j_1}, Y^{(d_2)} \leq y^{(d_2)}_{j_2} \,|\, x^{(2)}_{d_1}, x_{d_2}\right) = \int_{-\infty}^{\alpha^{d_1}_{j_{d_1}} - x^{(2)}_{d_1}\beta_{d_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(b - a\eta\right) d\eta, \tag{52}$$

where $a = \frac{\rho_{d_1,d_2}}{\sqrt{1-\rho^2_{d_1,d_2}}}$*,* $b = \frac{\alpha^{d_2}_{j_{d_2}} - x_{d_2}\beta_{d_2}}{\sqrt{1-\rho^2_{d_1,d_2}}}$*. Because* $\alpha^{d_1}_{j_{d_1}} - x^{(2)}_{d_1}\beta_{d_1} \leq 0$*, the right-hand side of (52) is strictly increasing in* a*. It is obviously also strictly increasing in* b*. This means that for any feasible* $a \in \mathbb{R}$ *we can find* $b_2(a)$ *such that*

$$P\left(Y^{(d_1)} \leq y^{(d_1)}_{j_1}, Y^{(d_2)} \leq y^{(d_2)}_{j_2} \,|\, x^{(2)}_{d_1}, x_{d_2}\right) = \int_{-\infty}^{\alpha^{d_1}_{j_{d_1}} - x^{(2)}_{d_1}\beta_{d_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(b_2(a) - a\eta\right) d\eta,$$

and $b_2(\cdot)$ *is a strictly decreasing function. Now consider the probability*

$$P\left(Y^{(d_1)} > y^{(d_1)}_{j_1}, Y^{(d_2)} \leq y^{(d_2)}_{j_2} \,|\, x^{(1)}_{d_1}, x_{d_2}\right) = \int_{\alpha^{d_1}_{j_{d_1}} - x^{(1)}_{d_1}\beta_{d_1}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(b - a\eta\right) d\eta,$$

where a *and* b *are the same as in (52). Because* $\alpha^{d_1}_{j_{d_1}} - x^{(1)}_{d_1}\beta_{d_1} \geq 0$*, the right-hand side of the last expression is strictly decreasing in* a*. It is obviously also strictly increasing in* b*. This means that for any feasible* $a \in \mathbb{R}$ *we can find* $b_1(a)$ *such that*

$$P\left(Y^{(d_1)} > y^{(d_1)}_{j_1}, Y^{(d_2)} \leq y^{(d_2)}_{j_2} \,|\, x^{(1)}_{d_1}, x_{d_2}\right) = \int_{\alpha^{d_1}_{j_{d_1}} - x^{(1)}_{d_1}\beta_{d_1}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(b_1(a) - a\eta\right) d\eta.$$

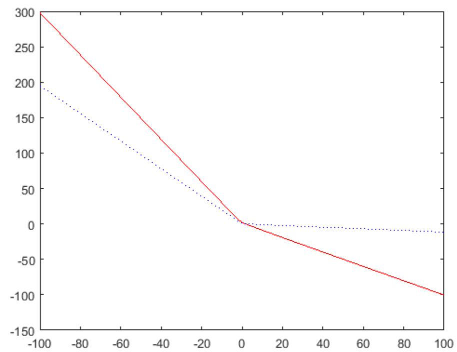
Note that since we only vary the first L_{d_1} *covariates in* x_{d_1}*, which are excluded from* x_{d_2}*, then* $alpha^{d_2}_{j_{d_2}} - x_{d_2}\beta_{d_2}$ *does not vary. This implies that* ρ_{d_1,d_2} *is identified because the strictly increasing* $b_1(\cdot)$ *and the strictly decreasing* $b_2(\cdot)$ *can intersect only once and the argument at that intersection is at* $\frac{\rho_{d_1,d_2}}{\sqrt{1-\rho^2_{d_1,d_2}}}$*, which can be inverted to give* ρ_{d_1,d_2}*.*

We now consider the case when both $\alpha^{d_1}_{j_{d_1}} - x^{(1)}_{d_1}\beta_{d_1}$ *and* $\alpha^{d_1}_{j_{d_1}} - x^{(2)}_{d_1}\beta_{d_1}$ *have the same sign. Suppose that they are both non-positive*[33] *Without a loss of generality,*

$$P\left(Y^{(d_1)} \leq y^{(d_1)}_{j_1}, Y^{(d_2)} \leq y^{(d_2)}_{j_2} \,|\, x^{(1)}_{d_1}, x_{d_2}\right) > P\left(Y^{(d_1)} \leq y^{(d_1)}_{j_1}, Y^{(d_2)} \leq y^{(d_2)}_{j_2} \,|\, x^{(2)}_{d_1}, x_{d_2}\right).$$

[33] If they are both non-negative, then instead of considering the conditional probabilities of $\{Y^{(d_1)} \leq y^{(d_1)}_{j_1}, Y^{(d_2)} \leq y^{(d_2)}_{j_2}\}$ we would consider the conditional probabilities of $\{Y^{(d_1)} > y^{(d_1)}_{j_1}, Y^{(d_2)} \leq y^{(d_2)}_{j_2}\}$.

FIGURE 22: Functions $b_2(\cdot)$ (solid line) and $b_1(\cdot)$ (dotted line)



Then both level functions $b_2(\cdot)$ and $b_1(\cdot)$ defined by equations

$$P\left(Y^{(d_1)} \leq y_{j_1}^{(d_1)}, Y^{(d_2)} \leq y_{j_2}^{(d_2)} \,|\, x_{d_1}^{(2)}, x_{d_2}\right) = \int_{-\infty}^{\alpha_{j_{d_1}}^{d_1} - x_{d_1}^{(2)}\beta_{d_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(b_1(a) - a\eta\right) d\eta$$

and

$$P\left(Y^{(d_1)} \leq y_{j_1}^{(d_1)}, Y^{(d_2)} \leq y_{j_2}^{(d_2)} \,|\, x_{d_1}^{(2)}, x_{d_2}\right) = \int_{-\infty}^{\alpha_{j_{d_1}}^{d_1} - x_{d_1}^{(1)}\beta_{d_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}} \Phi\left(b_2(a) - a\eta\right) d\eta$$

are strictly decreasing. However, the function $b_1(a)$ has a derivative that is strictly greater than the derivative of $b_2(a)$ for all a in the intersection of feasible sets. Moreover, for all low enough common feasible a the values of $b_1(a)$ are lower than the values of $b_2(a)$ and for all high enough a the values of $b_1(a)$ are higher than the values of $b_2(a)$. This situation is illustrated in Figure 22 which is obtained for specific realizations of Together with the strict inequality on the derivatives of these functions, these properties imply that these two functions may intersect only once. Their intersection is at $\frac{\rho_{d_1,d_2}}{\sqrt{1-\rho_{d_1,d_2}^2}}$, which can be inverted to give ρ_{d_1,d_2}. $\square$

To obtain a the identification of all the correlation coefficients in the multivariate normal distribution, one would verify that one of the conditions of Theorem 9 hold for each pair of dimensions (d_1, d_2). The fact that that correlations can be identified for each pair of dimensions at a time is a property of the lattice structure of the model. An interesting fact to note in Theorem 9 is that the identification of the correlation coefficients can be guaranteed even without the presence of exclusive covariates.

C.2 Non-lattice ordered probit

Here we do not present a set of clear-cut sufficient conditions that guarantee identification in a general non-lattice ordered probit model subject to Assumption 5. As discussed in Section 6, such conditions are difficult to derive (and even the tangentially related multinomial probit literature have not suggested such conditions). However, with the purpose of illustrating what kind of "sufficient variation" may be required we discuss identification in a bivariate non-lattice probit model. The results presented below rely on the presence of an exclusive covariate in at least one latent process. However, our simulations results in Section 8 indicate that this is not necessary.

Consider a special case of two dimensions and two ordered responses in each dimension. We outline the idea for identification when it is known that $\varepsilon = (\varepsilon_1, \varepsilon_2)'$ satisfies Assumption 5. In this 2×2 case the thresholds are $\alpha_{11}^{(1)}$, $\alpha_{12}^{(1)}$ and $\alpha_{11}^{(2)}$, $\alpha_{21}^{(2)}$. The fact that we have a coherent decision problem, or, in other words, our four rectangular regions partition the $\mathbb{R}^2$ plane implies that either $\alpha_{11}^{(1)} = \alpha_{12}^{(1)}$ or $\alpha_{11}^{(2)} = \alpha_{21}^{(2)}$ has to be satisfied. This is taken into account in Theorem 10 below. This theorem presents one set of sufficient conditions that guarantee identification.

Theorem 10 *Consider a bivariate ordered response model with two responses in each dimension. Suppose Assumption 5 holds.*

(a) (Sufficient variation in covariates in each dimension)

For dimension d, $d = 1, 2$, there are $k_d + 1$ points $\{x_d^{(i)}\}_{i=1}^{k_d+1}$ in $\mathcal{X}_d$ such that the matrix

$$\begin{pmatrix} 1 & x_d^{(1)} \\ 1 & x_d^{(2)} \\ \vdots & \vdots \\ 1 & x_d^{(k_d+1)} \end{pmatrix}$$

has rank $k_d + 1$.

(b) (Variation in x_1 or x_2; structure with $\alpha_{11}^{(2)} = \alpha_{21}^{(2)}$)

If $\alpha_{11}^{(1)} \neq \alpha_{12}^{(1)}$, then <u>either</u>

(b1) There is an exclusive covariate in x_1 with a non-zero coefficient. Without loss of generality, this variable is $x_{1,1}$ and its corresponding coefficient is $\beta_{1,1} \neq 0$. Also, there is x_2 in $\mathcal{X}_2$ that is observed with three different values of x_1 that differ only in $x_{1,1}$ – say, these are $x_1^{(i)} \equiv (x_{1,1}^{(i)}, x_{1,2:k_1})$, $i = 1, 2, 3,$, such that

$$P\left(Y^{(1)} = y_1^{(1)} \,|\, Y^{(2)} = y_1^{(2)}; x_1^{(i)}, x_2\right) \neq P\left(Y^{(1)} = y_1^{(1)} \,|\, Y^{(2)} = y_1^{(2)}; x_1^{(j)}, x_2\right) \quad i \neq j. \quad (53)$$

<u>or</u>

(b2) There is an exclusive covariate in x_2 with a non-zero coefficient. Without loss of generality, this variable is $x_{2,1}$ and its corresponding coefficient is $\beta_{2,1} \neq 0$. Also, there is $x_1 \in \mathcal{X}_1$

that is observed with two different values of x_2 that differ only in $x_{2,1}$ – say, these are $x_2^{(i)} \equiv (x_{2,1}^{(i)}, x_{2,2:k_2})$, $i = 1, 2,$, such that

$$\alpha_{21}^{(2)} - x_2^{(1)}\beta_2 \geq 0, \quad \alpha_{21}^{(2)} - x_2^{(2)}\beta_2 \leq 0.$$

(c) (Variation in x_1 or x_2; structure with $\alpha_{11}^{(1)} = \alpha_{12}^{(1)}$)

This is analogous to (b).

If $\alpha_{11}^{(2)} \neq \alpha_{21}^{(2)}$, then <u>either</u>

(c1) There is an exclusive covariate in x_2 with a non-zero coefficient. Without loss of generality, this variable is $x_{2,1}$ and its corresponding coefficient is $\beta_{2,1} \neq 0$. Also, there is $\tilde{x}_1$ in $\mathcal{X}_1$ that is observed with three different values of x_2 that differ only in $x_{2,1}$ – say, these are $\tilde{x}_2^{(i)} \equiv (\tilde{x}_{2,1}^{(i)}, x_{2,2:k_2})$, $i = 1, 2, 3,$, such that

$$P\left(Y^{(2)} = y_1^{(2)} \,|\, Y^{(1)} = y_1^{(1)}; \tilde{x}_1, \tilde{x}_2^{(i)}\right) \neq P\left(Y^{(2)} = y_1^{(2)} \mid Y^{(1)} = y_1^{(1)}; \tilde{x}_1, \tilde{x}_2^{(j)}\right) \quad i \neq j. \tag{54}$$

<u>or</u>

(c2) There is an exclusive covariate in x_1 with a non-zero coefficient. Without loss of generality, this variable is $x_{1,1}$ and its corresponding coefficient is $\beta_{1,1} \neq 0$. Also, there is $\tilde{x}_2 \in \mathcal{X}_2$ that is observed with two different values of x_1 that differ only in $x_{1,1}$ – say, these are $\tilde{x}_1^{(i)} \equiv (\tilde{x}_{1,1}^{(i)}, x_{1,2:k_1})$, $i = 1, 2,$, such that

$$\alpha_{21}^{(1)} - \tilde{x}_1^{(1)}\beta_1 \geq 0, \quad \alpha_{21}^{(1)} - \tilde{x}_1^{(2)}\beta_1 \leq 0.$$

(d) If $\alpha_{11}^{(1)} = \alpha_{12}^{(1)}$ and $\alpha_{11}^{(2)} = \alpha_{21}^{(2)}$, then at least one of the conditions of Theorem 9 is satisfied.

Then parameters β_1, β_2, $\alpha_{11}^{(1)}$, $\alpha_{12}^{(1)}$, $\alpha_{11}^{(2)}$, $\alpha_{21}^{(2)}$ and $\rho \equiv corr(\varepsilon_1, \varepsilon_2)$ are identified.

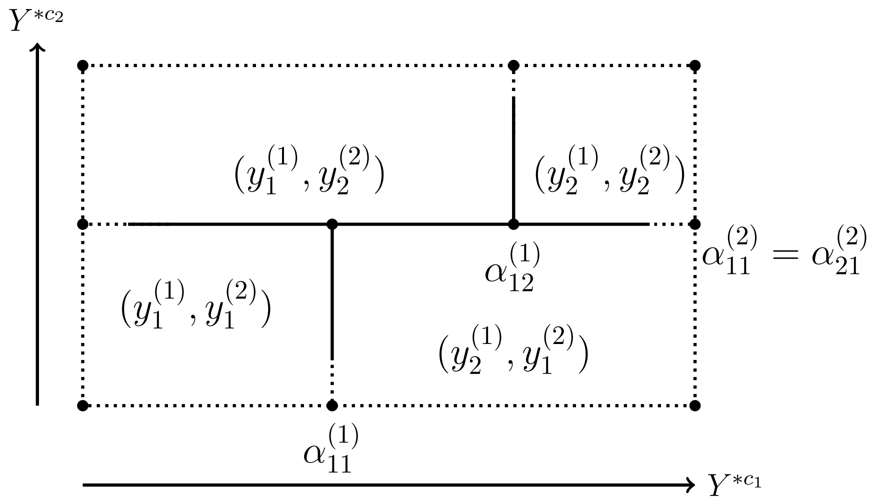
Proof of Theorem 10.

<u>Step 1</u>. Let us start by supposing that we know that in our threshold structure $\alpha_{11}^{(2)} = \alpha_{21}^{(2)}$ and denote this threshold as just $\alpha^{(2)}$, as depicted in Figure 23. We later discuss discuss how we move away from this supposition.

With a sufficient variation in x_2, we can identify parameters $\alpha^{(2)}$, β_2 simply because

$$P\left(Y^{(c2)} = y_1^{(2)} | x_2\right) = P\left(\varepsilon_2 \leq \alpha^{(2)} - x_2\beta_2 | x_2\right) = \Phi\left(\alpha^{(2)} - x_2\beta_2\right),$$

FIGURE 23: Step 1 of Figure 9



and hence,

$$\Phi^{-1}\left(P\left(Y^{(c2)} = y_1^{(2)}|x_2\right)\right) = \alpha^{(2)} - x_2\beta_2.$$

A sufficient variation in x_2 that ensures the identification of $\alpha^{(2)}$ and β_2 is guaranteed by condition (a).

Step 2. Now let us look at the identification of other parameters.

Step 2a. Suppose first that the condition (b1) is satisfied and take the x_2 that satisfies the property stated in that condition. Since $\alpha^{(2)}$ and β_2 are already identified, we know $q_2 \equiv \alpha^{(2)} - x_2\beta_2$. Suppose that $q_2 \leq 0$ and consider $P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \,|\, x_1^{(i)}, x_2\right)$, $i = 1, 2, 3$, where $x_1^{(i)}$ are chosen according to the condition (b1) as well. (If $q_2 > 0$ then instead we would consider the probabilities $P(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_2^{(2)} \,|\, x_1^{(i)}, x_2)$.)

Denote $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The Cholesky square root of Σ is $\Sigma^{\frac{1}{2}} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 \\ \rho & 1 \end{pmatrix}$ (so we have $\left(\Sigma^{\frac{1}{2}}\right)' \Sigma^{\frac{1}{2}} = \Sigma$). We have

$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \,|\, x_1^{(i)}, x_2\right) = P\left(\varepsilon_1 \leq \alpha_{11}^{(1)} - x_1^{(i)}\beta_1, \, \varepsilon_2 \leq q_2 | \, x_1^{(i)}, x_2\right).$$

This probability can be written as

$$P\left(\left(\Sigma^{\frac{1}{2}}\right)' \left(\Sigma^{-\frac{1}{2}}\right)' (\epsilon_1, \epsilon_2)' \leq (\alpha_{11}^{(1)} - x_1^{(i)}\beta_1, q_2)' \,|\, x_1^{(i)}, x_2\right). \tag{55}$$

Note that $\left(\Sigma^{-\frac{1}{2}}\right)' (\epsilon_1, \epsilon_2)'$ has the standard bivariate normal distribution and is independent of (x_1, x_2).

Denote such a standard bivariate random vector as $(\eta_1, \eta_2)'$. Then (55) can once again be rewritten as

$$P\left(\eta_1 \leq \frac{\alpha_{11}^{(1)} - x_1^{(i)}\beta_1}{\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}}\eta_2, \eta_2 \leq q_2 \,|\, x_1^{(i)}, x_2\right),$$

and further rewritten as

$$\int_{-\infty}^{q_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_2^2}{2}} \Phi\left(\frac{\alpha_{11}^{(1)} - x_1^{(i)}\beta_1 - \rho\eta_2}{\sqrt{1-\rho^2}}\right) d\eta_2.$$

Without loss of generality, we can suppose that in the condition (54)

$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \,|\, x_1^{(3)}, x_2\right) > P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \,|\, x_1^{(2)}, x_2\right) >$$
$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \,|\, x_1^{(1)}, x_2\right).$$

We will be able to obtain the identification of $\beta_{1,1}$, ρ and $q_{10} \equiv \alpha_{11}^{(1)} - x_1^{(1)}\beta_1$ if we identify $\delta_0 \equiv \frac{\rho}{\sqrt{1-\rho^2}}$ (strictly increasing in ρ), $\theta_0 \equiv \frac{q_{10}}{\sqrt{1-\rho^2}}$ and $\beta_{1,1}$. The identification will be obtained from the properties of the following function of two variables (δ, θ):

$$\psi(\delta, \theta) \equiv \int_{-\infty}^{q_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_2^2}{2}} \Phi(\theta - \delta\eta_2) d\eta_2. \tag{56}$$

Because $q_2 \leq 0$, the function $\psi(\cdot, \cdot)$ is strictly increasing in δ. Clearly, it is also strictly increasing in θ.

Thus, the identification of $\beta_{1,1}$, δ_0 and θ_0 will be shown if we establish that the following system of equations is solved by unique $\beta_{1,1}$, δ_0 and θ_0:

$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \,|\, x_1^{(h)}, x_2\right) = \psi(\theta_0 + \sqrt{1+\delta_0^2} \cdot \beta_{1,1}(x_{1,1}^{(1)} - x_{1,1}^{(h)}), \delta_0), \quad h = 1, 2, 3. \tag{57}$$

Condition (54) ensures that points (θ_0, δ_0), $(\theta_0 + \sqrt{1+\delta_0^2} \cdot \beta_{1,1}(x_{1,1}^{(1)} - w_{1,1}^{(2)}), \delta_0)$ and $(\theta_0 + \sqrt{1+\delta_0^2} \cdot \beta_{1,1}(w_{1,1}^{(1)} - w_{1,1}^{(3)}), \delta_0)$ lie on three different level curves of the function $\psi(\cdot, \cdot)$. Since function $\psi(\cdot, \cdot)$ is known and is strictly increasing in each variable, these level curves can be described as collections of points $(\upsilon_h(\delta), \delta)$ with a *known* and *strictly decreasing* $\upsilon_h(\cdot)$ defined on $\mathbb{R}$ (region for $\frac{\rho}{\sqrt{1-\rho^2}}$). The level curves υ_h, $h = 1, 2, 3$, are strictly ordered; their ordering and the sign of $\beta_{1,1}$ are immediately identified from the ordering of $P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \,|\, x_1^{(h)}, x_2\right)$, $h = 1, 2, 3$, given above.

Then the system (57) implies the following equations:

$$\upsilon_2(\delta_0) - \upsilon_1(\delta_0) = \sqrt{1+\delta_0^2} \cdot \beta_{1,1}(x_{1,1}^{(1)} - x_{1,1}^{(2)}) \tag{58}$$

$$\upsilon_3(\delta_0) - \upsilon_1(\delta_0) - \sqrt{1+\delta_0^2} \cdot \beta_{1,1}(x_{1,1}^{(2)} - x_{1,1}^{(3)}) = \sqrt{1+\delta_0^2} \cdot \beta_{1,1}(x_{1,1}^{(1)} - x_{1,1}^{(2)}). \tag{59}$$

Since $x_{1,1}^{(2)} \neq x_{1,1}^{(3)}$, these two equations capture different type of information. The uniqueness of $(\delta_0, \theta_0, \beta_{1,1})$ that solves this system (and, hence, the uniqueness of ρ, q_{10}) will be guaranteed by the properties of the level curves of function $\psi(\cdot,\cdot)$ formulated in Lemma 2 below.

Now, armed with the knowledge of ρ as well as all the parameters in the second dimension, we can go back to using the expression

$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \,|\, \tilde{x}_1, \tilde{x}_2\right) = \int_{-\infty}^{\alpha^{(2)} - \tilde{x}_2\beta_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_2^2}{2}} \Phi\left(\frac{\underbrace{\alpha_{11}^{(1)} - \tilde{x}_1\beta_1}_{\tilde{q}_{10}} - \rho\eta_2}{\sqrt{1-\rho^2}}\right) d\eta_2. \tag{60}$$

for any $(\tilde{x}_1, \tilde{x}_2) \in \mathcal{X}$. The integration limits in (60) are known. In fact, the only unknown if what we denoted as $\tilde{q}_{10}$. The right-hand side (60) is strictly increasing in $\tilde{q}_{10}$. Therefore, we can uniquely determine $\tilde{q}_{10}$ from the strict monotonicity of the right-hand side in $\tilde{q}_{10}$ and the knowledge of the left-hand side in (60).

Now, if we collect many of such points $\tilde{q}_{10}$ with enough variation in $\tilde{x}_1$, then it will be enough to uniquely determine parameters $\alpha_{11}^{(1)}$ and β_1. The condition for a sufficient variation in $\tilde{x}_1$ are given in part (a) of the theorem.

Now, in order to identify $\alpha_{12}^{(1)}$, consider

$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_2^{(2)} \,|\, \tilde{x}_1, \tilde{x}_2\right) = \int_{\alpha^{(2)} - \tilde{x}_2\beta_2}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_2^2}{2}} \Phi\left(\frac{\alpha_{12}^{(1)} - \tilde{x}_1\beta_1 - \rho\eta_2}{\sqrt{1-\rho^2}}\right) d\eta_2$$

for any $(\tilde{x}_1, \tilde{x}_2) \in \mathcal{X}$. Since $\alpha_{12}^{(1)}$ is the only unknown parameter on the right-hand side and the right-hand side is strictly monotone in $\alpha_{12}^{(1)}$, then $\alpha_{12}^{(1)}$ is identified in a straightforward way.

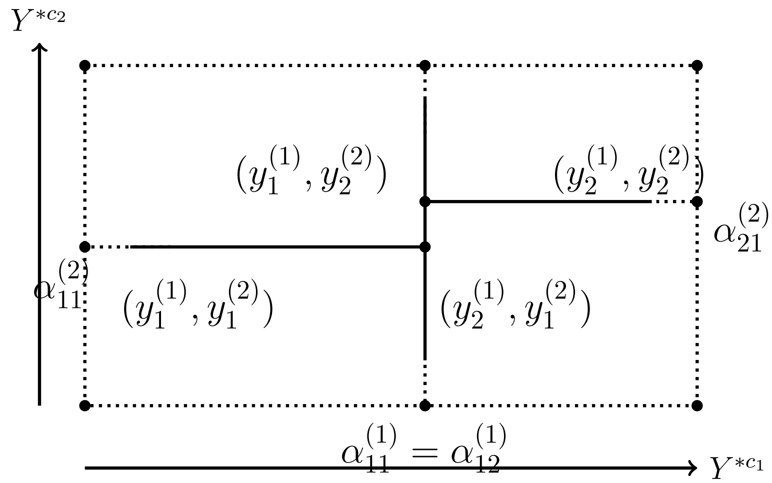
Step 2b). Suppose first that condition (b2) is satisfied and take the x_1 that satisfies the property stated in that condition. For $x_2^{(2)}$, consider the probability

$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_2^{(2)} \,|\, x_1, x_2^{(2)}\right) = \int_{-\infty}^{\alpha^{(2)} - x_2^{(2)}\beta_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_2^2}{2}} \Phi\left(\theta_0 - \delta_0\eta_2\right) d\eta_2, \tag{61}$$

where $\delta_0 = \frac{\rho}{\sqrt{1-\rho^2}}$, $\theta_0 = \frac{\alpha_{11}^{(1)} - x_1\beta_1}{\sqrt{1-\rho^2}}$. Because $\alpha^{(2)} - x_2^{(2)}\beta_2 \leq 0$, the right-hand side of (61) is strictly increasing in δ_0. It is obviously also strictly increasing in θ_0. This means that for any $\delta \in \mathbb{R}$ we can find $\theta_2(\delta)$ such that

$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_2^{(2)} \,|\, x_1, x_2^{(2)}\right) = \int_{-\infty}^{\alpha^{(2)} - x_2^{(2)}\beta_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_2^2}{2}} \Phi\left(\theta_2(\delta) - \delta\eta_2\right) d\eta_2,$$

Figure 24: Step 3 of Theorem 9



and $\theta_2(\cdot)$ is a strictly decreasing function.

For $x_2^{(1)}$, consider the probability

$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_2^{(2)} \,|\, x_1, x_2^{(1)}\right) = \int_{\alpha^{(2)} - x_2^{(1)}\beta_2}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_2^2}{2}} \Phi\left(\theta_0 - \delta_0 \eta_2\right) d\eta_2,$$

where δ_0 and θ_0 are the same as in (52). Because $\alpha^{(2)} - x_2^{(1)}\beta_2 \geq 0$, the right-hand side of the last expression is strictly decreasing in δ_0. It is obviously also strictly increasing in θ_0. This means that for any $\delta \in \mathbb{R}$ we can find $\theta_1(\delta)$ such that

$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_2^{(2)} \,|\, x_1, x_2^{(1)}\right) = \int_{-\infty}^{\alpha^{(2)} - x_2^{(1)}\beta_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_2^2}{2}} \Phi\left(\theta_1(\delta) - \delta \eta_2\right) d\eta_2,$$

and $\theta_1(\cdot)$ is a strictly increasing function. Note that since we only vary the exclusive covariate $x_{2,1}$ and, thus, $\alpha_{11}^{(1)} - x_1\beta_1$ does not vary, then ρ and $\alpha_{11}^{(1)} - x_1\beta_1$ are identified because the strictly increasing $\theta_1(\cdot)$ and the strictly decreasing $\theta_2(\cdot)$ can intersect only once.

Now, that the parameter ρ is identified, the identification of $\alpha_{11}^{(1)}$ and β_1 follows the same logic as in the case (b1) and obtained from a sufficient variation condition (a). The identification of $\alpha_{12}^{(1)}$ follows the same logic as in the case (b1).

Step 3.

The next question is whether can we distinguish this case from the case when in the first dimension the thresholds are the same and in the second dimension they are possibly different. The latter case is illustrated in Figure 24.

Suppose we already know that the model considered above (as pictured in Figure 23) is well-specified (consistent with the data). We want to show that if the alternative model in 24, where the thresholds in the first dimension are the same, is consistent with the data as well, then necessarily we have a model with a lattice structure. In other words, we can show that it is not possible for both models (in Figures 23 and 24) to be consistent with the data (distribution of observables) if at least one pair of thresholds in the same dimension contains distinct thresholds.

Suppose that it is possible for models of both types to rationalize the data with sets of parameters $(\beta_1, \beta_2, \alpha_{11}^{(1)}, \alpha_{12}^{(1)}, \alpha_{11}^{(2)}, \alpha_{21}^{(2)}, \rho)$ and $(\breve{\beta}_1, \breve{\beta}_2, \breve{\alpha}_{11}^{(1)}, \breve{\alpha}_{12}^{(1)}, \breve{\alpha}_{11}^{(2)}, \breve{\alpha}_{21}^{(2)}, \breve{\rho})$, respectively. Then the following equations are satisfied:

$$\Phi(\alpha^{(2)} - x_2\beta_2) = P\left(Y^{(c2)} = y_1^{(2)} \,|\, x_2\right) = \underbrace{P(x_2\breve{\beta}_2 + \varepsilon_2 \leq \breve{\alpha}_{11}^{(2)},\ x_1\breve{\beta}_1 + \varepsilon_1 \leq \breve{\alpha}^{(1)} | x_1, x_2)}_{P\left(Y^{(c1)}=y_1^{(1)},\, Y^{(c2)}=y_1^{(2)} \,|\, x_1, x_2\right)}$$

$$+ \underbrace{P(x_2\breve{\beta}_2 + \varepsilon_2 \leq \breve{\alpha}_{21}^{(2)},\ x_1\breve{\beta}_1 + \varepsilon_1 > \breve{\alpha}^{(1)} | x_1, x_2)}_{P\left(Y^{(c1)}=y_2^{(1)},\, Y^{(c2)}=y_1^{(2)} \,|\, x_1, x_2\right)}$$

$$= \begin{cases} P(x_2\breve{\beta}_2 + \varepsilon_2 \leq \breve{\alpha}_{21}^{(2)} | x_1, x_2) - P(\breve{\alpha}_{11}^{(2)} < x_2\breve{\beta}_2 + \varepsilon_2 \leq \breve{\alpha}_{21}^{(2)},\ x_1\breve{\beta}_1 + \varepsilon_1 \leq \breve{\alpha}^{(1)} | x_1, x_2), & \text{if } \breve{\alpha}_{21}^{(2)} > \breve{\alpha}_{11}^{(2)}, \\ P(x_2\breve{\beta}_2 + \varepsilon_2 \leq \breve{\alpha}_{11}^{(2)} | x_1, x_2) - P(\breve{\alpha}_{21}^{(2)} < x_2\breve{\beta}_2 + \varepsilon_2 \leq \breve{\alpha}_{11}^{(2)},\ x_1\breve{\beta}_1 + \varepsilon_1 \leq \breve{\alpha}^{(1)} | x_1, x_2), & \text{if } \breve{\alpha}_{11}^{(2)} \geq \breve{\alpha}_{221}^{(2)} \end{cases}$$

For illustration purposes suppose $\breve{\alpha}_{21}^{(2)} > \breve{\alpha}_{11}^{(2)}$ (the case of $\breve{\alpha}_{21}^{(2)} \leq \breve{\alpha}_{11}^{(2)}$ is considered analogously)). Then

$$\Phi(\alpha^{(2)} - x_2\beta_2) = \Phi(\breve{\alpha}_{21}^{(2)} - x_2\breve{\beta}_2)$$
$$- \int\int \phi(\eta_1, \eta_2)\, 1\left((-\infty, \breve{\alpha}_{11}^{(2)})' \leq \left(\breve{\Sigma}^{\frac{1}{2}}\right)' (\eta_1, \eta_2)' + (x_1\breve{\beta}_1, x_2\breve{\beta}_2)' < (\breve{\alpha}^{(1)}, \breve{\alpha}_{21}^{(2)})'\right) d\eta_1 d\eta_2$$
$$= \Phi(\breve{\alpha}_{21}^{(2)} - x_2\breve{\beta}_2)$$
$$- \int\int \phi(\eta_1, \eta_2)\, 1\left(\sqrt{1-\breve{\rho}^2}\eta_1 + \breve{\rho}\eta_2 < \breve{\alpha}^{(1)} - x_1\breve{\beta}_1, \breve{\alpha}_{11}^{(2)} < \eta_2 + x_2\breve{\beta}_2 \leq \breve{\alpha}_{21}^{(2)}\right) d\eta_1 d\eta_2,$$

where $\phi(\cdot, \cdot)$ is the density for the bivariate standard normal.

Suppose condition (b1) holds. Note that the fact that *both* structures are consistent with the observables and the fact that $\beta_{1,1} \neq 0$ will imply that $\breve{\beta}_{1,1} \neq 0$. Indeed, it is easy to see if e.g. we write the joint probability

$$P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \,|\, x_1, x_2\right) = \int_{-\infty}^{\alpha^{(2)} - x_2\beta_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_2^2}{2}} \Phi\left(\frac{\alpha_{11}^{(1)} - x_1\beta_1 - \rho\eta_2}{\sqrt{1-\rho^2}}\right) d\eta_2 \tag{62}$$

$$= \int_{-\infty}^{\breve{\alpha}_{11}^{(2)} - x_2\breve{\beta}_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta_2^2}{2}} \Phi\left(\frac{\breve{\alpha}^{(1)} - x_1\breve{\beta}_1 - \breve{\rho}\eta_2}{\sqrt{1-\breve{\rho}^2}}\right) d\eta_2, \tag{63}$$

and vary $x_{1,1}^{(h)}$ for instance by taking points $(x_1^{(h)}, x_Q)$, $h = 1, 2, 3$, that satisfy condition (54).

If we are in the situation of the lattice structure, then in the alternative model we have $\breve{\alpha}_{11}^{(2)} = \breve{\alpha}_{21}^{(2)}$ and, thus,

$$\Phi(\breve{\alpha}_{21}^{(2)} - x_2'\breve{\beta}_2) = \Phi(\alpha^{(2)} - x_2'\beta_2),$$

which further imply the equalities

$$\beta_2 = \breve{\beta}_2,\ \beta_1 = \breve{\beta}_1,\ \alpha_{11}^{(1)} = \alpha_{12}^{(1)} = \breve{\alpha}^{(1)},\ \alpha^{(2)} = \breve{\alpha}_{11}^{(2)} = \breve{\alpha}_{21}^{(2)} \tag{64}$$

for the parameters in the two models.

Suppose that $\breve{\alpha}_{11}^{(2)} \neq \breve{\alpha}_{21}^{(2)}$ and, thus, we are not dealing with the lattice structure. Then in the equation

$$\begin{aligned}\Phi(\alpha^{(2)} - x_2\beta_2) &= \Phi(\breve{\alpha}_{21}^{(2)} - x_2\breve{\beta}_2) \\ &\quad - \int\int \phi(\eta_1, \eta_2)\, 1\left(\sqrt{1-\breve{\rho}^2}\eta_1 + \breve{\rho}\eta_2 < \breve{\alpha}^{(1)} - x_1\breve{\beta}_1, \breve{\alpha}_{11}^{(2)} < \eta_2 + x_2\breve{\beta}_2 \leq \breve{\alpha}_{21}^{(2)}\right) d\eta_1 d\eta_2,\end{aligned} \tag{65}$$

the second term on the right-hand side depends on x_1.

As discussed above, $\breve{\beta}_{1,1} \neq 0$. Suppose for simplicity that $\breve{\beta}_{1,1} > 0$. We can consider (e.g. from condition (b1)) two points $\left(x_1^{(h)}, x_2\right)$, $h = 1, 2$, which differ only in the value of $x_{1,1}$. Without loss of generality suppose that $x_{1,1}^{(1)} > x_{1,1}^{(2)}$. Then under $\left(x_1^{(1)}, x_2\right)$ the region over which the integral on the right-hand side of (65) is calculated is strictly smaller than that under $\left(x_1^{(2)}, x_2\right)$. Therefore, under $(x_1^{(1)}, x_2)$ the right-hand side of (65) is strictly greater than that under $(x_1^{(2)}, x_2)$. However, the left-hand side remains the same under both $(x_1^{(h)}, x_2)$, $h = 1, 2$. This gives a contradiction meaning that the only situation in which both these competing models can rationalize the data is the case of the lattice structure and relations (64) hold.

Suppose condition (b2) holds.

To establish that we necessarily have $\breve{\alpha}_{11}^{(2)} = \breve{\alpha}_{21}^{(2)}$, instead of (65) we consider $\Phi(\breve{\alpha}^{(1)} - x_1\beta_1)$. Suppose for simplicity that $\alpha_{12}^{(1)} > \alpha_{11}^{(1)}$ (the case of $\alpha_{12}^{(1)} \leq \alpha_{11}^{(1)}$ is considered analogously)

$$\begin{aligned}\Phi(\breve{\alpha}^{(1)} - x_1\beta_1) &= \Phi(\alpha_{12}^{(1)} - x_1\beta_1) \\ &\quad - \int\int \phi(\eta_2, \eta_1)\, 1\left(\sqrt{1-\rho^2}\eta_2 + \rho\eta_1 < \alpha^{(2)} - x_2\beta_2, \alpha_{11}^{(1)} < \eta_1 + x_1\beta_1 \leq \alpha_{12}^{(1)}\right) d\eta_2 d\eta_1,\end{aligned} \tag{66}$$

implied by the fact that *both* structures are consistent with the observables and note the second term on the right-hand side depends on $x_{2,1}$ (since $\beta_{2,1} \neq 0$).

We can consider (e.g. from condition (b2)) two points $\left(x_1, x_2^{(h)}\right)$, $h = 1, 2$, which differ only in the value of $x_{2,1}$. Without loss of generality suppose that $x_{2,1}^{(1)} > x_{2,1}^{(2)}$ (and thus, $\beta_{2,1} > 0$). Then under $(x_1, x_2^{(1)})$ the region over which the integral on the right-hand side of (66) is calculated is strictly smaller than that under $(x_1, x_2^{(2)})$. Therefore, under $(x_1, x_2^{(1)})$ the right-hand side of (66) is strictly greater than that under $(x_1, x_2^{(2)})$. However, the left-hand side remains the same under both $(x_1, x_2^{(h)})$, $h = 1, 2$. This gives a contradiction meaning that the only situation in which both these competing models can rationalize the data is the case of the lattice structure and relations (64) hold.

Step 4. If we have a lattice structure meaning that $\alpha_{11}^{(2)} = \alpha_{21}^{(2)}$ and $\alpha_{11}^{(1)} = \alpha_{12}^{(1)}$, then all the parameters of the model, including the correlation ρ will be identified from conditions in Theorem 9. $\square$

Lemma 2 *Function $\psi(\cdot,\cdot)$ defined in (56) has level curves with the following property: For any three different level curves $(\upsilon_h(r), r)$, $h = 1, 2, 3$, any constant shift of the function $\frac{\upsilon_3(\delta)-\upsilon_1(\delta)}{\sqrt{1+\delta^2}}$ can intersect the function $\frac{\upsilon_2(\delta)-\upsilon_1(\delta)}{\sqrt{1+\delta^2}}$ at most once on $\mathbb{R}$.*

Remark 2 *If at least of the covariates in x_2 is exclusive and continuous and has a non-zero coefficient associated with it, then we can differentiate with respect to that covariate. Suppose it is $x_{2,1}$:*

$$\frac{\partial P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \mid x_1, x_2\right)}{\partial x_{2,1}} = -\beta_{2,1} \frac{1}{\sqrt{2\pi}} e^{-\frac{q_2^2}{2}} \Phi\left(\frac{q_{10} - \rho q_2}{\sqrt{1-\rho^2}}\right),$$

where $q_{10} \equiv \alpha_{11}^{(1)} - x_1\beta_1$. Since $\beta_{2,1}$ and q_2 are known, we know the left-hand side in the following expression:

$$\Phi^{-1}\left(-\frac{\sqrt{2\pi}}{\beta_{2,1}} \cdot e^{\frac{q_2^2}{2}} \cdot \frac{\partial P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \mid x_1, x_2\right)}{\partial x_{2,1}}\right) = \frac{q_{10} - \rho q_2}{\sqrt{1-\rho^2}}. \tag{67}$$

Take another value of $x_{2,1}^{(1)}$ and consider $x_2^{(1)} \equiv (x_{2,1}^{(1)}, x_{2,2:k_2})$ that differs from x_2 only in the value of the first component. Then analogously to above we know the left-hand side in the equation

$$\Phi^{-1}\left(-\frac{\sqrt{2\pi}}{\beta_{2,1}} \cdot e^{\frac{q_2^{(1)2}}{2}} \cdot \frac{\partial P\left(Y^{(c1)} = y_1^{(1)}, Y^{(c2)} = y_1^{(2)} \mid x_1, x_2\right)}{\partial x_{2,1}}\right) = \frac{q_{10} - \rho q_2^{(1)}}{\sqrt{1-\rho^2}}, \tag{68}$$

where $q_2^{(1)} \equiv \alpha^{(2)} - x_2\beta_2$.

It is easy to see now that ρ and q_{10} are identified from equations (67) and (68). If there are no continuous covariates among exclusive covariates in x_2, then instead of the partial derivatives we consider the differences.

D Additional examples, simulations, and results

D.1 Examples

This subsection contains three economics contexts that generate non-lattice models.

Example 2 (Empirically testing for selection in insurance markets) *Since the seminal work of Chiappori and Salanie (2000), several papers have empirically tested for asymmetric information in insurance markets by estimating the correlation in lattice bivariate ordered probit models. The two dependent variables are a dummy equal to one if an individual bought coverage (denoted y_1), and y_2, which is a discrete variable representing the number of potentially claimable events the individual is involved in (such as a crash in the case of autos, an illness with in the case of health etc.).[34] Chiappori and Salanie (2000) consider the correlation coefficient in a bivariate lattice probit using y_1 and y_2 as dependent variables, and a set of demographics and other regressors on the right-hand side.[35] A positive coefficient implies that those with private characteristics inducing coverage have private characteristics that increase the likelihood of a claimable event, implying some combination of adverse selection and moral hazard.[36] Evidence on the sign of the correlation across several papers is mixed. Chiappori and Salanie (2000) calculates a statistically insignificant coefficient of -0.02. Finkelstein and Poterba (2004) does not find adverse selection in coverage, but does find it on other dimensions of the contract. Finkelstein and McGarry (2006) finds no evidence of positive correlation between risk types and policy choices. Fang, Keane, and Silverman (2008) calculate a negative correlation coefficient and so find evidence of advantageous, rather than adverse selection in Medigap coverage. Cohen (2005) finds adverse selection in auto insurance choices. Taking the validity of the correlation coefficient in the bivariate lattice probit model as given, the mixed evidence on selection in insurance markets ran contrary to the intuition of theorists discussing insurance markets prior to the empirical work. However, it is important to consider that data do not exist on markets that are missing (or have unravelled) because of particularly strong adverse selection – the data from non-missing markets are a selected set.*

Despite this, the mixed evidence may result from the incorrect use of a bivariate probit model in this context. If moral hazard exists, then the thresholds that determine the number of claimable insurance events (α_2) will depend on the presence or absence of coverage (y_1), implying a non-lattice model. Resultantly, researchers interested in empirically testing for asymmetric information through the correlation

[34] Typically authors use a dummy for y_2, equal to one if the individual has any accident or crash etc. .

[35] Chiappori and Salanie (2000) also suggests to calculate the correlation between the generalized residuals from two univariate probits, and provides a nonparametric χ^2 test.

[36] Some papers estimated the correlation in a context where one of selection or moral hazard was not possible. Otherwise, most papers attempting to separate selection from moral hazard either require exogenous variation in coverage assignment or a structural model.

of a bivariate model should use non-lattice models.

Example 3 (Advertisement spillover effects) *Consider three firms (A, B and C) making choices about advertisement. Advertisement slots come in discrete packages with quality (or impact) Q_h and price P_h, $h = 1, \ldots, H$.*[37] *Each company buys at most one advertisement package. Bundles are ordered so that $Q_{h+1} > Q_h$ and $P_{h+1} > P_h$. The marginal price of quality must be nondecreasing in the level of quality so that*

$$\frac{Q_{h+1} - Q_h}{P_{h+1} - P_h}$$

is increasing in h. There are positive spillovers from advertisement. These spillovers have a triangular structure: A's advertisement affects B's and C's profitability, and B's advertisement affects C's profitability

$$\begin{aligned}
U_A(I_A, Q_{h_A}) &= I_A - P_{h_A} + \tau_A Q_{h_A}, \\
U_B(I_B, Q_{h_B}, Q_{h_A}) &= I_B - P_{h_B} + \tau_B Q_{h_B}(1 + \phi_{AB}(Q_{h_A})), \\
U_C(I_C, Q_{h_C}, Q_{h_B}, Q_{h_A}) &= I_C - P_{h_C} + \tau_C Q_{h_C}(1 + \phi_{AC}(Q_{h_A}))(1 + \phi_{BC}(Q_{hB})),
\end{aligned}$$

where I_ℓ denotes the profitability of firm ℓ in the absence of advertisement, known nonnegative functions $\phi_{AB}(\cdot)$, $\phi_{AC}(\cdot)$, $\phi_{BC}(\cdot)$ capture the spillover effect from rivals' advertisements and τ_ℓ stands for firm ℓ's marginal valuation of advertisement. Assume

$$\tau_\ell = x\beta_\ell + \varepsilon_\ell,$$

with observed x_ℓ and unobserved ε_ℓ. In the equilibrium A, B and C choose Q_{h_A}, Q_{h_B} and Q_H, respectively, if and only if

$$\frac{P_{h_A+1} - P_{h_A}}{Q_{h_A+1} - Q_{h_A}} < \tau_A \leq \frac{P_{h_A+2} - P_{h_A+1}}{Q_{h_A+2} - Q_{h_A+1}}$$

$$\frac{P_{h_B+1} - P_{h_B}}{(Q_{h_B+1} - Q_{h_B})\,(1 + \phi_{AB}(Q_{h_A}))} < \tau_B \leq \frac{P_{h_B+2} - P_{h_B+1}}{(Q_{h_B+2} - Q_{h_B+1})\,(1 + \phi_{AB}(Q_{h_A}))}$$

$$\frac{P_{h_C+1} - P_{h_C}}{(Q_{h_C+1} - Q_{h_C})\,(1 + \phi_{AC}(Q_{h_A}))\,(1 + \phi_{BC}(Q_{h_B}))} < \tau_C \leq \frac{P_{h_C+1} - P_{h_C}}{(Q_{h_C+1} - Q_{h_C})\,(1 + \phi_{AC}(Q_{h_A}))\,(1 + \phi_{BC}(Q_{h_B}))}.$$

Thus, this system leads to a hierarchical ordered response model. We can think of A first determining all the decisions rules (thresholds) for herself, then B determining all the decisions rules (thresholds) for herself given the decision rules by A, and finally C determining all the decisions rules (thresholds)

[37] *For example, newspaper adverts are discrete in the sense that there may be a finite set of pages and sizes available. The nearer the advert is to the front and the larger is the size, the higher is the quality.*

for herself given the decision rules by A and B.

In the absence of spillover effects or if such spillovers were additive rather than multiplicative, we would end up with a lattice ordered response model.

Examples 1 and 3 are special cases of strategic interactions models that result in coherent non-lattice frameworks. Example 2 relates to economic decisions made in several dimensions by a *single* agent where actions taken by her in one dimension affect the payoff in the the other dimension and, thus, result in a non-lattice structure. This is exactly our main intended framework. We can give additional examples with that single decision maker paradigm:

1. A decision maker considers buying good A without knowing how valuable good B will be but knows good B is more/less enjoyable if they have good A.
2. An academic is deciding whether to work on paper A, with the idea in mind to do paper B. The success of paper B is unknown but will be more substantial if paper A is a success.
3. An inventor is deciding to patent invention A, knowing that patenting invention A will improve the success of invention B, but not yet knowing whether invention B will work or not.
4. A political party is deciding whether to spend money at the start of their period of leadership, knowing that this could help them at the end of their tenure but that they might not need to do it if their ratings are sufficiently high.
5. A high school graduate is deciding whether to take training / do degree A, knowing that they will face a choice of doing job B (perhaps taking over the family business). They don't know the success of taking job B, but they know that choice A will affect it.

There are other formats of simultaneous equations that result in non-lattice models. We finish this section with a final example of this.

Example 4 (Financial transfers and distress) *In this example, there are two dimensions with three ordered responses in each dimension. One dimension corresponds to a parent company and the other to a subsidiary. Let Y_p^* and Y_s^* stand for continuous metrics of financial distress of these companies before any financial transfers between companies. The financial distress of one company (either parent or subsidiary) when the other company is financially healthy may necessitate financial transfers from the latter to the former. We can also expect that when both companies are financially distressed, the extent of mutual help may be more limited. Also, it is possible that a moderately financially distressed subsidiary may have a better chance of getting financial support from the parent company than a* *severely distressed subsidiary, as in the latter case, the parent company may give up on the subsidiary. To summarize, various cases of financial distress are possible, and depending on the case, different mutual transfer scenarios will realize. Let* H_p^* *and* H_s^* *denote the latent financial distress post-transfers and* $\pi > 0$ *denote weights. Then we model* H_p^* *and* H_s^* *as*

$$
\begin{aligned}
H_p^* &= Y_p^* \cdot \left(\sum_{j_p=1}^3 \sum_{j_s=1}^3 \pi_{j_p,j_s}^{(p)} 1 \left(\alpha_{j_p-1,j_s}^{(p)} < Y_p^* \leq \alpha_{j_p,j_s}^{(p)}, \alpha_{j_p,j_s-1}^{(s)} < Y_s^* \leq \alpha_{j_p,j_s}^{(s)} \right) \right), \\
H_s^* &= Y_s^* \cdot \left(\sum_{j_p=1}^3 \sum_{j_s=1}^3 \pi_{j_p,j_s}^{(s)} 1 \left(\alpha_{j_p-1,j_s}^{(p)} < Y_p^* \leq \alpha_{j_p,j_s}^{(p)}, \alpha_{j_p,j_s-1}^{(s)} < Y_s^* \leq \alpha_{j_p,j_s}^{(s)} \right) \right),
\end{aligned}
$$

$\alpha_{0,j_s}^{(p)} = -\infty$, $\alpha_{j_p,0}^{(s)} = -\infty$, $\alpha_{0,3}^{(p)} = \infty$, $\alpha_{3,0}^{(s)} = \infty$, *and the thresholds* $\alpha_{j_p,j_s}^{(p)}$ *and* $\alpha_{j_p,j_s}^{(s)}$, $j_p = 1, 2$, $j_s = 1, 2$, *split the plane into a non-lattice structure. Moreover, suppose that* $\pi_{j_p,j_s}^{(p)} \cdot \alpha_{j_p,j_s}^{(p)}$ *does not depend on* s *and is monotonic in* j_p, *and also* $\pi_{j_p,j_s}^{(s)} \cdot \alpha_{j_p,j_s}^{(s)}$ *does not depend on* p *and is monotonic in* j_s. *In this specification, weights* $\pi_{j_p,j_s}^{(p)} > 0$ *and* $\pi_{j_p,j_s}^{(s)} > 0$ *already incorporate the impact of mutual financial transfers and* $\alpha_{j_p,j_s}^{(p)}$, $\alpha_{j_p,j_s}^{(s)}$ *capture various ranges of pre-transfers financial distress for both companies.*

We can take $Y_p^* = \alpha_p + x_p'\beta_p + \varepsilon_p$ *and* $Y_s^* = \alpha_s + x_s'\beta_s + \varepsilon_s$, *where* x_p *and* x_s *include various financial indicators of the parent and subsidiary, respectively. There are three discrete measures* Y_p *and* Y_s *of financial distress post-transfers denoted 0, 1 and 2, with 0 representing that a company is healthy, 1 representing moderate financial distress and 2 representing severe financial distress. The discrete outcomes are determined according to the univariate ordered response models*

$$
\begin{aligned}
Y_p = j_p - 1 &\iff h_{j_p-1}^{(p)} < H_p^* \leq h_{j_p}^{(p)}, \quad j_p = 1, 2, 3, \\
Y_s = j_s - 1 &\iff h_{j_s-1}^{(s)} < H_s^* \leq h_{j_s}^{(s)}, \quad j_s = 1, 2, 3,
\end{aligned}
$$

where $h_0^{(p)} = h_0^{(s)} = -\infty$, $h_3^{(p)} = h_3^{(s)} = \infty$, $h_{j_p}^{(p)} = \alpha_{j_p,j_s}^{(p)} \cdot \pi_{j_p,j_s}^{(p)}$ *and* $h_{j_s}^{(s)} = c_{j_p,j_s}^{(s)} \cdot \pi_{j_p,j_s}^{(s)}$. *This results in the non-lattice model*

$$
Y_p = j_p - 1, \;\; Y_s = j_s - 1 \quad \iff \quad \alpha_{j_p-1,j_s}^{(p)} < Y_p^* \leq \alpha_{j_p,j_s}^{(p)}, \;\; \alpha_{j_p,j_s-1}^{(s)} < Y_s^* \leq \alpha_{j_p,j_s}^{(s)},
$$

which maps continuous processes for financial distress before transfers into discrete measures of financial distress post-transfers.

D.2 Simulations

Design 2: additional details

First we present a figure of the true latent variable space in the 4 x 3 model and then we provide a table with the simulation means and standard deviations of thresholds.

Figure 25: Latent variable space for two equations: design 2

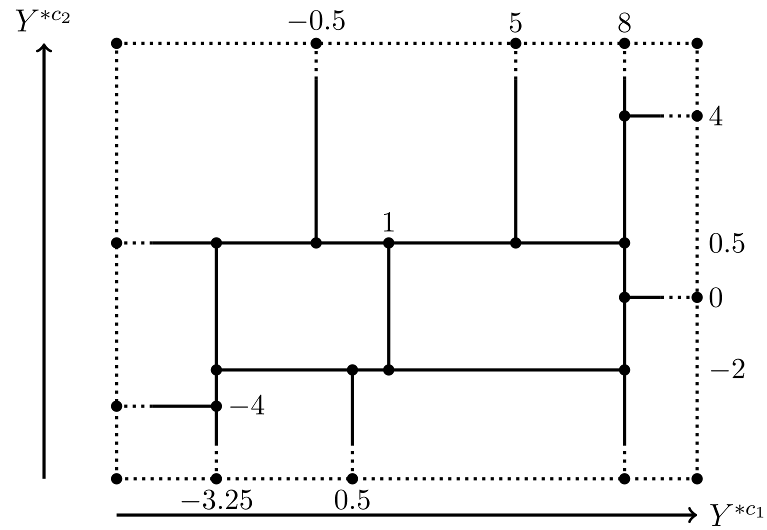


Table 3: Simulation results design 2: thresholds

Parameter	Truth	Non-lattice model	Lattice model
$\alpha_{10}^{(1)}$	-3.25	-3.27 (0.12)	-1.48 (0.04)
$\alpha_{11}^{(1)}$		-3.24 (0.12)	
$\alpha_{12}^{(1)}$	-0.5	-0.50 (0.07)	
$\alpha_{20}^{(1)}$	0.5	0.51 (0.09)	1.59 (0.04)
$\alpha_{21}^{(1)}$	1	0.97 (0.14)	
$\alpha_{22}^{(1)}$	5	5.02 (0.13)	
$\alpha_{30}^{(1)}$	8	8.03 (0.19)	5.12 (0.09)
$\alpha_{31}^{(1)}$		8.03 (0.19)	
$\alpha_{32}^{(1)}$		8.03 (0.19)	
$\alpha_{01}^{(2)}$	-4	-3.94 (0.32)	-1.10 (0.04)
$\alpha_{11}^{(2)}$	-2	-2.04 (0.16)	
$\alpha_{21}^{(2)}$		-1.99 (0.09)	
$\alpha_{31}^{(2)}$	0	-0.01 (0.09)	
$\alpha_{02}^{(2)}$	0.5	0.50 (0.05)	0.90 (0.04)
$\alpha_{12}^{(2)}$		0.50 (0.05)	
$\alpha_{22}^{(2)}$		0.50 (0.05)	
$\alpha_{32}^{(2)}$	4	3.99 (0.17)	

Notes: Table 3 reports the sample mean and sample standard deviations (in parentheses) of the estimates of the design 2 threshold parameters, over 250 repeated samples. The "Non-lattice model" column provide estimates from using the newly proposed non-lattice bivariate ordered probit model. The "Lattice model" column assumes a lattice structure on the latent variable space.

Table 4: Simulation results design 3

Parameter	**Truth**	**Non-lattice model**	**Lattice model**
β_1	1.5	1.50 (0.03)	0.73 (0.02)
γ_1	-4	-4.01 (0.07)	-2.10 (0.04)
β_2	3	3.03 (0.13)	0.48 (0.02)
γ_2	-6	-6.06 (0.25)	-1.29 (0.04)
δ_2	1	1.01 (0.05)	0.22 (0.01)
ρ	0.5	0.51 (0.07)	-0.87 (0.01)

Notes: Table 4 reports the sample mean and sample standard deviations (in parentheses) of the estimates of the model parameters, over 250 repeated samples. See table 1 notes for further details about the columns.

Design 3: 7×2

We consider a design that creates a 7×2 non-lattice structure on the latent variable space. Figure 26 illustrates the non-lattice structure in D. We run this simulation to showcase the ability of our method to arbitrarily extend the number of values taken by the discrete variables.

In this design, the common regressor x is drawn from uniform $[-2, 2]$ and both latent equations have excluded regressors $w_1, w_2 \overset{iid}{\sim} t_5$. We also include an additional regressor z_2 in equation 2, drawn from a logistic (2,1) distribution. The parameter corresponding to z_2 is denoted δ_2, so that the latent equations read

$$
\begin{aligned}
Y^{*c_1} &= x\beta_1 + w_1\gamma_1 + \varepsilon_1 \\
Y^{*c_2} &= x\beta_2 + w_2\gamma_2 + \varepsilon_2 + z_2\delta_2
\end{aligned}
$$

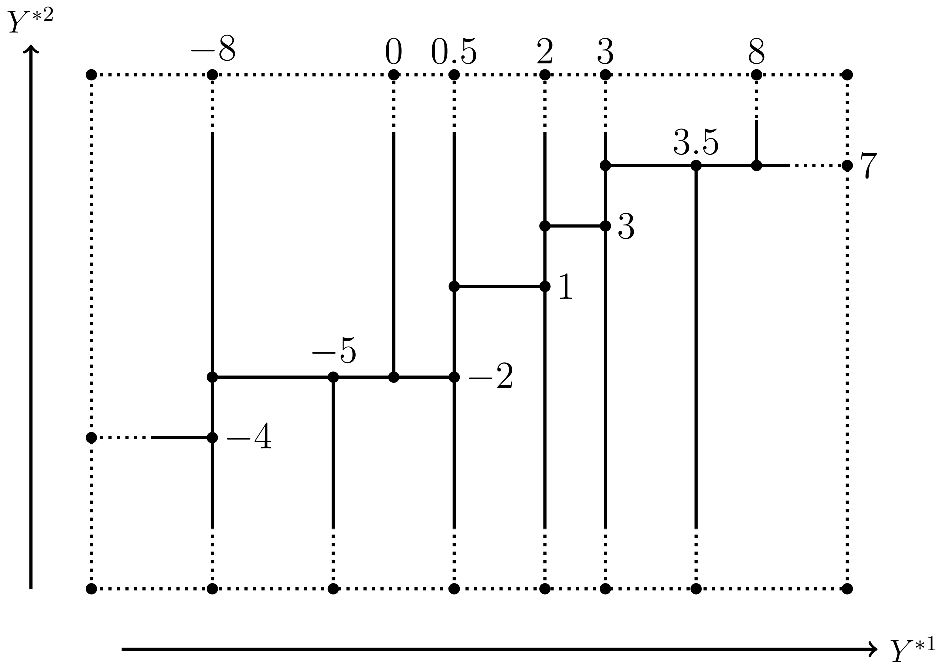
The parameter values β_1, β_2, γ_1 and δ are the same as in design 2, and $\gamma_2 = -6$, $\delta_2 = 1$. Table 5, found in appendix D, provides the values and simulation results for the thresholds. Table 4 presents the results for the regression parameters and the correlation coefficient. The finite sample bias in the newly proposed method is far smaller than existing methods, and again the bivariate lattice ordered probit method cannot estimate ρ with any degree of accuracy.

Table 5: Simulation results design 3: thresholds

Parameter	**Truth**	**Non-lattice model**	**Lattice model**
$\alpha_{10}^{(1)}$	-8	-8.02 (0.15)	-4.62 (0.08)
$\alpha_{11}^{(1)}$		-8.012 (0.15)	
$\alpha_{20}^{(1)}$	-5	-5.00 (0.11)	-0.99 (0.03)
$\alpha_{21}^{(1)}$	0	0.00 (0.04)	
$\alpha_{30}^{(1)}$	0.5	0.50 (0.04)	0.13 (0.02)
$\alpha_{31}^{(1)}$		0.50 (0.04)	
$\alpha_{40}^{(1)}$	2	2.01 (0.05)	1.18 (0.03)
$\alpha_{41}^{(1)}$		2.001 (0.05)	
$\alpha_{50}^{(1)}$	3	3.01 (0.06)	1.89 (0.04)
$\alpha_{51}^{(1)}$		3.01 (0.06)	
$\alpha_{60}^{(1)}$	3.5	3.51 (0.07)	2.85 (0.05)
$\alpha_{61}^{(1)}$	8	8.03 (0.17)	
$\alpha_{01}^{(2)}$	-4	-4.03 (0.25)	0.20 (0.03)
$\alpha_{11}^{(2)}$	-2	-2.01 (0.12)	
$\alpha_{21}^{(2)}$	-2	-2.01 (0.12)	
$\alpha_{31}^{(2)}$	1	1.01 (0.15)	
$\alpha_{41}^{(2)}$	3	3.03 (0.21)	
$\alpha_{51}^{(2)}$	7	7.08 (0.29)	
$\alpha_{61}^{(2)}$	7	7.08 (0.29)	

Notes: Table 5 reports the sample mean and sample standard deviations (in parentheses) of the estimates of the design 3 parameters, over 250 repeated samples. The "Non-lattice model" column provide estimates from using the newly proposed nonlattice bivariate ordered probit model. The "Lattice model" column assumes a lattice structure on the latent variable space.

FIGURE 26: Latent variable space for two equations: Design 3



D.3 Applications

D.3.1 Link back to the first application in Section 9.

Table 6: Estimation coefficients: identity theft and cash opinion

Variable	**Probit**	**O-probit**	**Non-lattice**	**Lattice**
Identity theft				
Low Income	-0.17 (0.04)		-0.12 (0.11)	-0.17 (0.04)
Age	0.01 (0.00)		0.01 (0.00)	0.01 (0.00)
Male	0.02 (0.04)		0.05 (0.07)	0.02 (0.04)
Low Education	-0.24 (0.05)		-0.17 (0.12)	-0.24 (0.05)
Opinion on cash				
Low Income		0.17 (0.03)	0.14 (0.07)	0.17 (0.03)
Age		0.002 (0.00)	0.002 (0.00)	0.002 (0.00)
Male		0.12 (0.03)	0.12 (0.05)	0.12 (0.03)
Low Education		0.18 (0.04)	-0.16 (0.08)	0.18 (0.04)
ρ	NA	NA	0.48 (0.66)	-0.04 (0.02)
N	4633	4633	4633	4633

Notes: Table 6 reports coefficient estimates from the identity theft and cash opinion specification. Columns labelled "Probit" provides estimates from univariate ordered probit models. The "Non-lattice" column provide estimates from using the newly proposed non-lattice bivariate ordered probit model. The "Lattice" column assumes a lattice structure, but estimates the two equations jointly. Standard errors are reported in parentheses, and are typical standard errors except for the Non-lattice model where they are bootstrapped.

D.3.2 Link back to the second application in Section 9.

TABLE 7: Estimation coefficients: bitcoin familiarity and optimism

Variable	**O-Probit**	**O-Probit**	**Non-lattice**	**Lattice**
Familiarity with Bitcoin				
LOW INCOME	-0.16 (0.06)		-0.11 (0.05)	-0.16 (0.06)
AGE	-0.02 (0.00)		-0.02 (0.00)	-0.01 (0.00)
MALE	0.55 (0.06)		0.42 (0.06)	0.55 (0.06)
LOW EDUCATION	-0.54 (0.09)		-0.40 (0.09)	-0.54 (0.09)
Bitcoin "optimism"				
LOW INCOME		0.07 (0.06)	-0.00 (0.06)	0.07 (0.06)
AGE		-0.01 (0.00)	-0.01 (0.00)	-0.01 (0.00)
MALE		-0.13 (0.05)	0.02 (0.06)	-0.13 (0.05)
LOW EDUCATION		0.13 (0.07)	-0.00 (0.08)	0.13 (0.07)
ρ	NA	NA	0.84 (0.23)	0.03 (0.03)
N	1818	1818	1818	1818

Notes: Table 7 reports coefficient estimates from the cryptocurrency specification. See notes in Table 6 for details on the columns. Standard errors are reported in parentheses.

D.3.3 Additional application: Adoption of online payment instruments

We study the degree of broad bracketing in the adoption of online payment instruments. We entertain the idea that individuals may decide jointly which online payment instruments to adopt. Two of the leading modern payment methods are PayPal and Google Pay. The relationship between the adoption of these two online payment instruments is not immediate. Consider an individual who learns about the existence of online payment methods. Three possibilities immediately come to mind. First, the individual may choose *between* two payment methods, favoring the adoption of a single online payment device. Choosing between payment methods would imply that the payment methods are substitutes. Second, since there are some options to synchronize PayPal and Google Pay accounts, there may be synergies, and as a result, they may be complements. Third, the individual may narrowly bracket and decide whether to adopt each payment method independently, unaware of any relationship. These three cases would each imply a different threshold structure and estimating a non-lattice model will differentiate among them.

We estimate the following model

$$\begin{aligned} \textsc{PayPal} &= x_1\beta_1 + \varepsilon_1 \\ \textsc{GooglePay} &= x_2\beta_2 + \varepsilon_2 \end{aligned}$$

where PAYPAL and GOOGLEPAY are dummies equal to 1 if the individual uses PayPal or Google Pay to make a purchase or pay another person in the last year, respectively. The covariate vectors x_1 and x_2 are identical and are teh same as in the two applications.[38]

Table 8 displays estimates of β and the correlation parameter ρ across lattice and non-lattice specifications. The β coefficients are broadly similar across estimation methods, but there is a significant difference between the estimated value of ρ. In the lattice model, the estimate is 0.30, whereas in the non-lattice model, it is 0.80.

Figures 27 and 28 show the estimated thresholds across estimation methods. In the non-lattice model, the value of $\alpha_{12}^{(1)}$ is much larger than $\alpha_{11}^{(1)}$. This difference in thresholds implies that individuals consider both mobile payment options when deciding which to adopt, suggesting some broad bracketing in this decision.[39] More specifically, individuals' utility from PayPal needs to pass a much higher threshold to lead to adoption if the individual already has Google Pay relative to if they don't. Hence, the two options are substitutes as opposed to complements. The lattice model has no way of allowing for this complementarity/substitutability. Instead, it forces that individual decision structures on the adoption of PayPal and Google Pay to be independent, consistent with narrow bracketing.

[38] More specifically, LOW INCOME is 1 if annual household income falls below $50,000, and LOW EDUCATION is 1 if the individual did not attend college.

[39] We can reject the null of equality of thresholds at 5% significance.

FIGURE 27: Estimates from the payment instrument example, assuming a lattice model

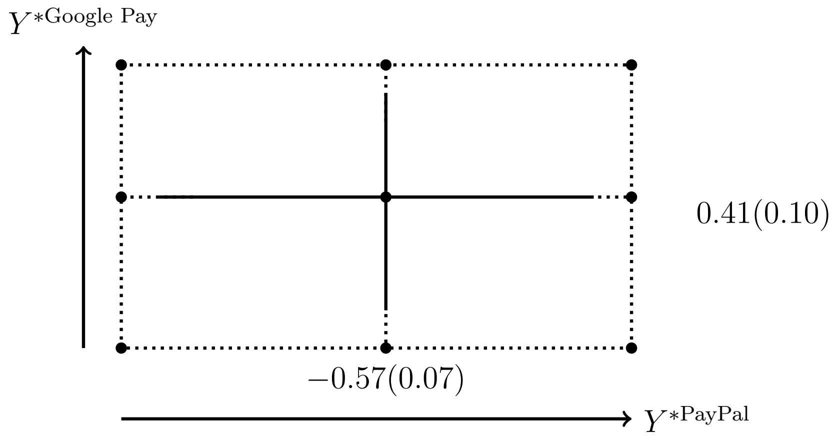


FIGURE 28: Estimates from the payment instrument example, assuming a non-lattice model

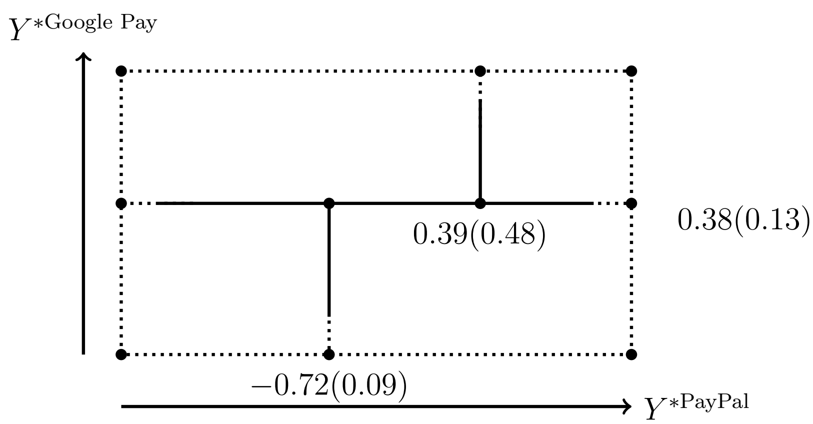


TABLE 8: Estimation coefficients: online payment instruments

Variable	**Probit**	**Probit**	**Non-lattice**	**Lattice**
PayPal adoption				
Low Income	-0.31 (0.04)		-0.28 (0.04)	-0.31 (0.04)
Age	-0.01 (0.00)		-0.01 (0.00)	-0.01 (0.00)
Male	-0.04 (0.04)		-0.02 (0.04)	-0.04 (0.04)
Low Education	-0.34 (0.05)		-0.34 (0.04)	-0.34 (0.05)
Google Pay adoption				
Low Income		-0.01 (0.06)	-0.02 (0.08)	-0.01 (0.06)
Age		-0.02 (0.00)	-0.02 (0.00)	-0.02 (0.00)
Male		0.11 (0.06)	0.07 (0.07)	0.10 (0.06)
Low Education		-0.17 (0.08)	-0.15 (0.09)	-0.17 (0.07)
ρ	NA	NA	0.80 (0.30)	0.30 (0.03)
N	4634	4634	4634	4634

Notes: Table 8 reports coefficient estimates from the PayPal and Google Pay specification.